TRADABLE PERMIT MARKETS FOR THE CONTROL OF POINT AND NONPOINT SOURCES OF WATER POLLUTION: TECHNOLOGY-BASED V. COLLECTIVE PERFORMANCE-BASED APPROACHES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of

Philosophy in the Graduate School of The Ohio State University

By

Michael A. Taylor, B.A., M.S.

*****

The Ohio State University
2003

Dissertation Committee:

Professor Alan Randall, Advisor

Professor Brent L. Sohngen

Professor Donald C. Hubin

Professor Earl Epstein

Approved By:

_____

Advisor
Interdisciplinary Program

ABSTRACT

The United States Environmental Protection Agency has begun to encourage innovative market-based approaches to address nonpoint source water pollution. These water quality trading programs have the potential to achieve environmental standards at a lower overall cost. Two fundamental questions must be answered before these benefits can be realized: How will trades between point and nonpoint sources be monitored and enforced?; and, How will nonpoint sources be included within a trading market?

Point-nonpoint source trading can be accommodated through either a technology-based or performance-based approach. The technology-based approach accommodates trading through the use of a proxy for unobservable, individual nonpoint source emission reductions. While trading ratios can effectively deal with the uncertainty associated with using a proxy for actual abatement, they are inefficient and ineffective for dealing with problems of hidden action. The alternative use of performance-based trading approaches requires the use of team contracts that provide individual incentives linked to the performance of the entire group. Such contracts must be designed to overcome both adverse selection and moral hazard problems. Performance-based approaches promise efficiency gains in terms of reducing the problems of asymmetric information, and by

introducing flexibility into the choice of nonpoint source abatement technologies and practices.

Nonpoint sources are exempted from direct regulation under the polluter-pays-principle. As a result, their participation in trading markets is voluntary, thus preventing a baseline cap on pre-trade emissions. To determine whether this arrangement should be changed, we must ask if there something that morally prohibits the direct regulation of nonpoint sources of pollution. While a morally relevant distinction can be made between point and nonpoint sources of emission based on differences in the ability to observe individual emission levels, this relevance is limited to the case of performance-based policy instruments. The moral legitimacy of applying the polluter-pays-principle to nonpoint sources of pollution must be made on a case by case basis, as it is dependent upon existing social, economic, and other practical factors. However, it can be stated that there is no general moral barrier to prohibit the application of the polluter-pays-principle to nonpoint sources of pollution.

Dedicated to Lynn, Duncan, and Nicolas

ACKNOWLEDGMENTS

This dissertation owes many thanks to numerous people for their assistance, patience, and kindness. Without their support, it would have never been completed. First, I am indebted to my graduate committee members. My faculty advisor, Alan Randall, has been a mentor, teacher, and friend throughout this process. From the outset he has been supportive and encouraging of my desire to pursue a graduate degree that combined the fields of economics and philosophy. He helped me appreciate not only how to do research, but how to do it well. Brent Sohngen and Don Hubin provided invaluable assistance in refining and focusing the scope of this project into a manageable and meaningful effort. Earl Epstein also provided his enthusiasm and useful insights. Finally, although not formal committee members, the input of Steven Wu and Helen Pushkarskaya contributed substantially to the quality and content.

This dissertation research benefited from the financial support of various institutions. I would like to thank Resources for the Future for their support through the Joseph L. Fisher Dissertation Fellowship. In addition, the USDA provided research grant funding through the National Research Initiative Grant Program. I must also thank the Ohio State University Environmental Policy Initiative for employing me during the majority of my time at Ohio State.

While I have been at Ohio State, numerous fellow graduate students have come and gone leaving a piece of themselves with me. Todd Gabe, Suzie Greenhalgh, Jeff Hopkins, Maria Pagura, Mike Pruit, and Mehnaz Safavian have all educated, inspired, and amused me, although not necessarily in that order. The friendship and support of Diana Lantz has been greatly appreciated.

Last, but not least, I must thank my family. To my brother John, I express my gratitude for all of the rush-hour phone calls. To my mom Natalie and my dad Robert, I owe my most sincere thanks for instilling within me the belief that with enough hard work I could achieve whatever goals I desired. I am truly grateful for the loving support and patience of my wife Lynn and our boys, Duncan and Nicolas, without which I would never have been able to dedicate the time and effort required to write this dissertation. Yes, my paper is finally done.

VITA

September 13, 1966 …	Born – Cheshire, CT
1988 ……...…	B.A. Economics, University of Connecticut
1991 – 1993 ……	Graduate Teaching and Research Associate, Department of Applied Economics, Virginia Tech
1993 ..……..	M.S. Applied Economics, Virginia Tech
1995 – Present ...…….	Ph.D. One-of-a-Kind Program, The Ohio State University
1996 – 1998 ……	Graduate Teaching and Research Associate, Department of Agricultural, Environmental, and Development Economics, The Ohio State University
1996 – 2001 ……	Instructor, Department of Equine Science Department of Life Sciences Otterbein College
1998 – 1999 ……	Joseph L. Fisher Dissertation Fellow Resources for the Future, Washington, DC
1998 – Present ……	Graduate Administrative Associate, The Environmental Policy Initiative, The Ohio State University

PUBLICATIONS

1. Randall, A, and MA Taylor, "Incentive-Based Solutions to Agricultural Environmental Problems: Recent Developments in Theory and Practice" *Journal of Agricultural and Applied Economics.*, 32(2), 221 (2001)

FIELDS OF STUDY

Major Fields: Interdisciplinary Program: Agricultural, Environmental, and Development Economics; Philosophy; Environmental Science

TABLE OF CONTENTS

Abstract ii

Acknowledgments v

Vita vii

List of Figures x

List of Tables xi

Chapters:

1. Introduction 12

1.1 A Primer on the Economic Theory of Tradable Permit Markets 14
1.1.1 Structuring and Allocating Rights to Pollute 16
1.1.2 Trading Rights 20
1.2 Difficulties Incorporating Nonpoint Sources 21
1.2.1 Nonpoint Source Participation 22
1.2.2 Basis of Exchange 23
1.3 Improving the Prospects for Point-Nonpoint Pollution Trading 26

2. Technology-based Trading Approach 30

2.1 A Model of Technology-based Trading Under Symmetric Information . 33
2.2 The Safety Rule Approach 36
2.3 Linking the Safety Rule and the Trading Ratio 47
2.4 Should the Optimal Trading Ratio Always be Less Than One? 49
2.5 Two Reasons the Optimal Trading Ratio May Be Greater Than One 53
2.6 A Model of Technology-based Trading Under Asymmetric Information 56
2.7 Conclusions 62

3. Collective Performance-based Trading Approach 64

3.1 Introduction 66

3.2 Existing Literature on Collective Monitoring and Enforcement 68
3.3 A Model of a Two-stage Team Contract for Water Quality Trading 70
3.4 A Budget-breaking Team Contract to Avoid Moral Hazard in Teams... 74
3.5 An Auction Mechanism to Determine Team Entry and Sharing Rules.. 90
3.6 Uniform Price Team Entry Auction..... 93
3.7 Discriminating Price Team Entry Auction 97
3.8 Summary and Conclusions 104

4. What is Morally Special About Nonpoint Source Pollution?..... 106

4.1 Harm, Regulation, and the Polluter-pays-principle 107
4.2 The Moral Relevance of the Point/Nonpoint Distinction 112
4.2.1 Differential Uncertainty 114
4.2.2 Differential Observability 116
4.3 Moral Relevance of “Differential Observability”: Technology-based . 117
4.4 Moral Relevance of “Differential Observability": Performance-based 118
4.4.1 Performance-based Standard Setting 119
4.4.2 Performance-based Standard Enforcement..... 120
4.5 Collective Penalties and Imposing Costs on the Innocent..... 121
4.6 Justifying the Use of Collective Penalties 124
4.6.1 A Sketch of a PPP Collective Performance-based Program..... 126
4.6.2 A Sketch of a Collective Subsidy-based Program 126
4.6.3 Consequentialist Concerns..... 128
4.6.4 Retributivist Concerns 130
4.7 Some Practical Factors That Should Be Considered 131
4.8 Summary and Conclusions 133

5. Conclusions..... 135

Appendicies:

A. Model Notation 145

B. Shortfall Abatement Strategy 146

C. Nonparticipating Abatement Strategy 150

D. Uniform Price Auction (Shortfall Strategy)..... 153

E. Discriminating Price Auction (Shortfall Strategy) 156

Bibliography 161

LIST OF FIGURES

Figure Page

2.1: Risk Neutral Safety Rule for Stochastic Emissions 38
2.1: Risk Neutral Safety Rule for Stochastic Emissions 38
2.2: Risk Averse Safety Rule for Stochastic Emissions 39
2.3: Stochastic Damages Loss and Gain –Linear Environmental Damage Curve 42
2.4: Stochastic Damages Loss and Gain – Convex Environmental Damage Curve 44
2.5: Safety Rule for Stochastic Damages – Convex Environmental Damage Curve 46
2.6: Optimal Trading Ratio When Trading Decreases the Variance of Total Emissions . 52
2.7: The Mean Effect of Hidden Action on the Optimal Trading Ratio 61
3.1: First-Best Allocation of Contract Price and Quantity 83
3.2: Second-Best Allocation of Contract Quantity and Price (Surplus Strategy) 87
3.3: Second-Best Allocation of Contract Quantity and Price (Shortfall Strategy) 88
3.4: Optimal Bid Price and Quantity Uniform Price Auction (Surplus) 98
3.5: Optimal Bid Price and Quantity Discriminating Price Auction (Surplus) 102
D.1: Optimal Bid Price and Quantity Uniform Price Auction (Shortfall) 155
E.1: Optimal Bid Price and Quantity Discriminating Price Auction (Shortfall) 160

LIST OF TABLES

Table Page

3.1 Weather/Payment Contingent Scenarios..... 76

CHAPTER 1

INTRODUCTION

More than thirty years after its inception, the Clean Water Act can at best be categorized as a guarded success. Undeniably, there are success stories, with Lake Erie and the Cuyahoga River serving as excellent local examples. Lake Erie has been resuscitated after being declared dead twenty-years ago, and the infamous Cuyahoga is no longer prone to combustion. However, achievements are often overshadowed by the myriad of persistent and emerging water quality problems.

The majority of the accomplishments of the Clean Water Act can be attributed to the successful control of point source emissions. The Act's policy for nonpoint sources of emissions, such as urban stormwater runoff, agricultural runoff, and atmospheric deposition, has not been as effective. Approximately forty percent of all sampled waterways currently fail to meet their designated water quality standard, which results in more than 75 percent of the population living within ten miles of a polluted waterway (EPA, 2003). Nonpoint source emissions and agricultural emissions in particular, are the most significant contributors to current water quality problems in the United States (EPA, 2003). Agricultural emissions are the leading source of impairment in the nation's rivers and lakes (Water Quality Inventory). In addition, nutrient and sediment loading from

agriculture is the most significant contributor to emerging water quality problems, such as hypoxia in the Gulf of Mexico, and decreased fish populations in the Chesapeake Bay.

The financial scale of nonpoint source pollution programs is quite large. Section 319 of the 1986 reauthorization of the Clean Water Act, for example, authorized the EPA to spend approximately $130 million annually on nonpoint source pollution programs (Sohngen and Taylor, 1998). Between 1996 and 2002, the U.S. Department of Agriculture distributed over $1.6 billion to farmers through the Environmental Quality Incentive Program, and the Wildlife Habitat Incentive Program, and additional funding was provided to maintain 36 million acres of land under contract in the Conservation Reserve Program (Sohngen and Taylor, 1998). In addition to the federal cost-sharing programs, many states are also allocating large sums of funding to voluntary nonpoint source water quality programs. The Natureworks program in Ohio, for example, provides $1.5 million per year for different watershed projects that install abatement technologies, particularly for riparian zone enhancement (Sohngen and Taylor, 1998). Therefore, both the effects of nonpoint source emissions, and the regulatory expenditures to control them, are quite significant.

As a result, the EPA is encouraging innovative solutions, such as water quality trading, to facilitate bringing nonpoint sources under the regulatory umbrella. The implementation of market-based mechanisms akin to pollution permit trading programs has been suggested for broad range of environmental concerns, including air pollution control, wetlands and shoreline mitigation, habitat protection, and resource compensation. However, the use of trading programs for water quality control has been

receiving the most attention in recent years. By January 2003, at least thirty-seven water quality trading programs were in the developmental or implementation stage in the United States (EPA, 2003). "The United States Environmental Protection Agency ("EPA") believes that market-based approaches to water quality trading provide greater flexibility and have the potential to achieve water quality and environmental benefits greater than would otherwise be achieved under more traditional regulatory approaches" (EPA, 2003).

1.1 A Primer on the Economic Theory of Tradable Permit Markets

Economists have long argued for policy instruments that take maximum advantage of voluntary exchange, with its efficiency and Pareto-safety properties. As early as the 1960's, Crocker (1966) and Dales (1968a, 1968b) were arguing for the establishment of markets in the right to pollute, as a means of controlling air and water pollution. This is accomplished by establishing private property rights (in terms of allowable discharges) within a public goods setting. Tradable permit markets promise efficiency in terms of achieving stated environmental quality at the lowest cost. Efficiency is served by allowing individuals to transfer their right to a permitted level of emissions to other polluters. Polluters with low abatement costs will reduce emissions below their permitted level, and then sell the excess credit to an individual with higher abatement costs. Thus, the market encourages the rights to pollute to be distributed to the highest valued users.

There are two types of efficiency effects associated with tradable permit markets: static and dynamic. The static effects are realized through the switch in regulatory focus from command-and-control to emissions. Allowing individuals to choose their abatement practices provides flexibility in meeting regulatory goals. Managers are able to select those abatement technologies and practices that best suit their unique needs. This provides the polluter with the means of selecting the lowest-cost method for reducing the required level of emissions. The dynamic effect has also been called the induced innovation effect. Making the right to pollute transferable, provides surplus allowances with a market value. This encourages innovation to find low-cost abatement means to reduce emissions beyond one's permitted level.

While the institutional structure of permit markets can be quite diverse, all are developed in three distinct stages. The first is clearly defining the overall environmental quality goals of the program. The next two stages include allocating the right to pollute, and allowing the tranfer of rights. I will briefly discuss each of these stages in terms of the general theory. The extension of tradable permit markets to include nonpoint sources will then be discussed in terms of some difficulties encountered at each stage.

Analysis that links desired ambient conditions with some measure of total allowable emissions is required. This has been called the "bubble" concept in air pollution markets, as it determines the ambient threshold for emissions of an entire airshed, as if it had a large bubble over the market's entire geographic region. This

emphasizes one of the main points of all tradable permit markets: it doesn't matter who reduces pollution, as long as it is reduced.[1]

In markets for water pollution the same concept can be thought of as a "bowl". Many factors play in the determination of the appropriate scale and distance in water pollution trading markets. Environmental performance typically has a strong regional dimension. The one organizing principle for all water pollution trading is the incontrovertible fact that water tends to flow downhill. Thus the size and shape of the "bowl" is determined in direct reference to the hydrological boundaries which define watersheds and catchments.

The determination of both current and total allowable levels of emissions is required in order to give a baseline from which to measure environmental quality gains. In addition, this information determines the individual emissions that will be allocated among the pollution sources, as the sum of the individual emissions standard must sum to the total allowable level to ensure both market efficiency and effectiveness. This process is greatly facilitated by the EPA's current requirement for Total Maximum Daily Load analysis in impaired watersheds.

1.1.1 Structuring and Allocating Rights to Pollute

Market transactions are usually discussed in terms of the exchange of physical objects, or *property*. However, the substance of every market is not the exchange of *objects*, but of *rights*, and therefore should not be examined in terms of an exchange of

[1] This statement has a caveat in terms of spatial considerations for non-uniformly mixing pollutants, and this will be discussed later in the chapter.

property, but in terms of the exchange of *property rights*. Property rights establish the basis for economic activity, because they designate the relationship between people and things. A tradable permit is a bundle of property rights, which entitles the holder to discharge a designated amount of a pollutant under specific restrictions.

Ambient permits and emission permits were the first types of tradable permits discussed by Crocker (1966) and Dales (1968a, 1968b), and later formalized by Montgomery (1972). An emission permit confers the right to discharge pollutants at a specified rate, while an ambient permit confers the right to discharge pollutants in terms of concentrations at various receptor areas (Montgomery, 1972). A trading system based on ambient permits requires the polluter to hold a portfolio of permits, since its discharge has varying effects on each receptor area. The emission permit requires only a single permit market because it has only a single receptor area. Tradable permit markets based on either method have been shown to achieve a minimum-cost equilibrium, provided that certain conditions are met (Montgomery, 1972; Tietenberg, 1985; Atkinson and Tietenberg, 1982; Krupnick, et al., 1983; McGartland and Oates, 1985). In particular, the sum of allocated permits must equal the allowable quantity of discharge.

Later additions to the tradable permit taxonomy include pollution offsets, and modified pollution offsets. Pollution offsets (Krupnick, et al, 1983) accomplish the same goal as ambient permits, but do not require dischargers to hold multiple market permits. Instead, a trading ratio that incorporates the relative contribution of each discharger to each receptor area is utilized. Modified pollution offsets introduced by McGartland and Oates (1985) redefine water quality standards at each receptor area as being equal to

either the regulatory standard or the pre-standard water quality level, whichever is lower. Therefore, water quality degradation even below the regulatory standard is not permitted.

The diversity of tradable permit rights structures available to the regulator allow tradable permit markets to accommodate a variety of pollution types. Pollutants are categorized based on their temporal (stock v. flow) and spatial (uniformly mixing v. non-uniformly mixing) characteristics. The type of pollution being regulated greatly affects the choice of tradable permit system employed.

The difference between stock and flow pollutants is based on their ability to be assimilated into the natural environment. Stock pollutants, such as heavy metals and toxics, are poorly assimilated and emissions in the current time period will have continuing effects into the future. Flow pollutants, such as nitrogen and phosphorous, are quickly assimilated and do not have intertemporal effects. The temporal characteristics of a pollutant will determine if rights must be based on both the quantity and timing of discharges. Most trading markets have concentrated on assimilative pollutants.

A pollutant's mixing characteristic (uniform or non-uniform) determines the spatial effects of its discharge. The ambient concentration of a uniformly mixing pollutant depends upon the total emissions of all dischargers, but not on the distribution of these dischargers (Tietenberg, 1985). A non-uniformly mixing pollutant is one that has localized effects. The ambient concentration depends on both total emissions and the location of dischargers. A uniformly mixing pollutant only requires a single receptor area to monitor the ambient level of pollution in the watershed. A non-uniformly mixing pollutant requires multiple receptor areas to account for localized ambient concentrations.

In the case of a watershed requiring a single receptor area, all of the tradable permit markets perform identically (McGartland, 1988). However, in the presence of multiple receptor areas, the choice of tradable permit system can involve tradeoffs between administrative transaction costs and market transaction costs.

Water quality trading markets are primarily focused on the reduction of nutrients, such as nitrogen and phosphorous. These are uniformly mixing, flow pollutants. This provides the regulator with broad discretion in developing the type of emission right that can be incorporated within the market.

Once the type of pollution right is determined, the regulator must decide on how to distribute the rights among polluters. Within competitive tradable permit markets, the method chosen to initially distribute permits among dischargers does not affect the efficiency of the market. The choice is a matter of equity and political feasibility. One method for initial allocation is through a sales process. Permits are allocated to those dischargers who value them most, because all dischargers must purchase their way into the market. Single price auctions, discriminating markets, and unrestricted markets are all valid sales methods. Revenues generated through sales methods can be used to offset the start-up and operating costs of the tradable permit system, because they result in a large transfer of funds from polluters to the control agency (Eheart, et al., 1981). Sales methods are not likely to gain enthusiastic support from the market participants, however, and this can be a major impediment to the adoption of voluntary WQT markets.

Free distribution is more desirable from the viewpoint of the market participants, but it requires the consideration of many equity issues. Free distribution methods result

in some individuals receiving objects of value from the government at no cost. Therefore, the question of who should be considered a market participant is of greater importance in this type of distribution scheme. How the tradable permits are allocated can be based on any criteria including, but not limited to, pollutant throughput, population served, and total property value served (Eheart, et al., 1981). However, all of these criteria can be attacked as being arbitrary. The best method may be to base the initial allocations on an amended list of existing emissions, or "grandfathering" (Noll, 1982)

1.1.2 Trading Rights

The structure of the market, which determines how and when rights will be redistributed or exchanged, is the critical element in determining whether equilibrium will be achieved in practice (Hahn, 1983). Various methods have been proposed, ranging from unrestricted market exchange to quarterly single price auctions. The choice of the mechanism for exchange is made based on the regulator's need to monitor and record all trades. Each trade must be tracked in order to assure compliance with the regulatory standards. In addition, the maintenance of a database of potential buyers and sellers, and current market prices, reduces transaction costs for market participants. In active markets with large numbers of dischargers, the administrative costs of tracking a market with free exchange may become prohibitively high. In such situations, an auction mechanism may be a more feasible method of exchange. The majority of the literature assumes that

dischargers are able to buy and sell permits with relative ease at competitive prices and with incidental transaction costs, whenever benefits of trade exist (Malik, 1990).

In the case of WQT, the size of the market is limited by the physical boundaries of the watershed. In most cases, WQT markets are focused on at the smaller sub-watershed level. This reduces the transaction problems inherent in larger permit trading markets, such as the national SO_2 market.

1.2 Difficulties Incorporating Nonpoint Sources within Tradable Permit Markets

While there is considerable literature addressing permit market design, relatively little of it deals with extending permit markets to include nonpoint sources. It is often argued that including nonpoint source pollution within a permit trading market is difficult because the direct monitoring of individual emission is impractical. To compound this problem, emissions are affected by random weather events, and uncertainty exists regarding the effectiveness of pollution abatement controls. The degree to which individual nonpoint source emissions can be practically observed is a muddled issue within the current economics literature. I will contend throughout this dissertation that while direct monitoring of individual nonpoint sources is impractical, there is a great deal of accessible indirect information that allows for reliable approximation of actual individual emissions and abatement. This leads to some required modifications of the traditional permit market design to allow for the inclusion of nonpoint sources of pollution.

1.2.1 Nonpoint Source Participation

The introduction of nonpoint sources into the trading market can occur in either one of two ways. One way is to bring nonpoint sources under the same regulatory umbrella as point sources. In this case, the nonpoint sources would be subject to the polluter-pays-principle, and directly regulated through individual emissions standards that limit the amount effluent in their runoff. Then point and nonpoint sources would be allowed to exchange their rights, with the lowest cost sources selling excess credits to the higher cost sources. This would represent a dramatic departure from the current treatment of nonpoint sources under the Clean Water Act. Nonpoint sources do not participate in the Clean Water Act's primary regulatory mechanism, the National Pollutant Discharge Elimination System. Instead, efforts to control nonpoint source emissions, agricultural nonpoint sources in particular, are dealt with through voluntary cost-sharing programs that are subsidized through public funds.

Alternatively, the nonpoint sources can be invited into the tradable permit system on a voluntary basis. This has been the method of choice in all water quality trading programs to date. In this arrangement, point sources induce nonpoint control through partial or full subsidization of the nonpoint source's abatement costs. In this case, the incentive for nonpoint source participation is profit driven, and depends upon the cost of nonpoint source control being less expensive than further point source control. Nonpoint sources are not full participants in the current water quality trading markets. Their only role is as potential suppliers of pollution abatement. They do not need to purchase abatement from other sources because their own emissions are not regulated.

One criticism of the current voluntary participation arrangement is that it fails to produce an enforceable baseline from which to begin the trading process (Malik, et al. 1994). This is the basis of cap-and-trade markets, such as the SO_2 market. An enforceable baseline ensures the regulator that any voluntarily adopted reductions in emissions will not be offset by new or increased nonpoint source emissions in the watershed. To establish a cap-and-trade water quality market, the establishment of an enforceable nonpoint source baseline would require a roll back of the exemption to the polluter-pays-principle.

1.2.2 Basis of Exchange

Since nonpoint source emissions are difficult to observe at the individual level, existing trading programs have resorted to monitoring abatement technology (e.g., best management practices) rather than performance. As a result, trade within current and proposed water quality markets involves heterogeneous goods (point source discharges and nonpoint source best management practices).

Trading ratios have been introduced to allow for the exchange of heterogeneous goods within a tradable permit market (Mendelsohn, 1986; Hahn, 1989). The trading ratio specifies the number of units of nonpoint pollution reduction, estimated by modeling the effectiveness of the chosen abatement technology, that must be exchanged for a single unit of point source pollution. The optimal trading ratio depends on the expected performance of nonpoint source abatement technology as well as the

uncertainty associated with nonpoint source emissions. This uncertainty has two sources. The first derives from the weather driven nature of nonpoint source pollution. The second is the considerable uncertainty regarding the effectiveness of nonpoint source pollution abatement controls. A trading ratio greater than one provides a safety margin for the environment. By requiring more than one unit of (estimated) nonpoint source reduction credit, deviations from the expected abatement performance of the abatement technologies are less likely to result in violations of the regulatory standards. To assure that regulatory goals are met, the trading ratio tends be set cautiously high, but high trading ratios impede trading, thus undermining the *raison d'etre* for permit trading.

The current technology-based water quality trading markets have shown little success to date in terms of trading[2]. The two oldest markets, Dillon Reservoir, CO and Tar-Pamlico River Basin, NC, have both been in operation since the 1980's but have yet to see any voluntary trading between point and nonpoint sources. One possible explanation is the stylized fact, discussed above, that efficiency gains from the switch to performance standards for point sources have reduced the demand for permits. This trend was documented in the early stages of the $S0_2$ tradable permit market (Burtraw, 1997). However, before we accept this explanation too complacently, we should entertain a second possibility: trading markets that have been introduced are too restrictive, and too many bureaucratic controls remain, so that permit exchange is impeded by market design.

[2] This is not be meant to imply that the markets have not been successful in improving water quality. In both cases the water quality has been improved beyond the initial water quality standards. In both cases the switch to performance standards led to previously unexpected low-cost alternatives for the point sources, and in the Tar-Pamlico River Basin there was substantial trading between point sources (Woodward, 2001).

An alternative to monitoring abatement technology is to monitor the nonpoint sources on the basis of performance. The promise of this approach is that it removes the regulator's uncertainty about the effectiveness of the nonpoint source abatement technologies. This eliminates the need for a trading ratio, which provides an increase in the potential of trading. Uncertainty as to the effectiveness of on-farm abatement technology is borne by the farmers (who are best able to handle it), allowing greater efficiency in the permit market as regulators and point source polluters enjoy a higher level of certainty. All parties would gain from the increased efficiency of permit markets.

If it is conceded that monitoring individual nonpoint sources on the basis of performance is technically difficult, and thus likely to be prohibitively expensive, arrangements based on collective monitoring at the sub-watershed level might be considered. Monitoring ambient water quality at the sub-watershed level is a simple process. The mapping of ambient water quality into collective emissions can be quite accurate when considering relatively small geographic regions, such as sub-watersheds and sub-catchments. The assumption of an accurate mapping of ambient water quality into collective emissions is made throughout the dissertation. The difficulty is to provide the right incentives to individual nonpoint sources to link observations of collective performance to appropriate individual actions. Griffin and Bromley (1982) have suggested the use of estimated individual nonpoint discharges, derived from the monitoring of total loadings in the catchment, determined through a biophysical model relating inputs to loadings and ambient water quality standards. Segerson (1990) has suggested liability bonding. The game theory literature offers "scapegoat" and

“massacre” solutions (Rasmussen, 1987), variations on the theme that all firms will make appropriate abatement effort if collective performance is monitored and randomly chosen individuals punished in the event the collective target is not met. Schemes for punishment of all members of a group for shortfalls in collective performance, which avoid the arbitrariness of “scapegoat” and “massacre”, have also been suggested (Segerson, 1988).

Further research is required to refine methods for enforcing performance standards via collective monitoring of nonpoint sources within a water quality trading market. There are two key requirements for an acceptable collective monitoring and enforcement mechanism. The first is that it transmits to group members clear and readily comprehensible incentives that are consistent with group goals. In effect, incentive-compatibility and simplicity (which may come into conflict) are valued. The second requirement is that penalties and rewards imposed on individuals do not violate ordinary notions of fairness. Collectively imposed penalties and random penalties for a group shortfall may be considered unfair to those group members who did not shirk.

1.3 Improving the Prospects for Point-Nonpoint Pollution Trading

The rationale behind the research in this dissertation is that market-based water pollution regulation systems may provide cost savings by bringing relatively low cost nonpoint source abatement into the regulatory mix. This proposition will be investigated in the context of water quality trading markets that include both point and nonpoint

sources. These types of markets have received increased attention recently as regulators search for more cost-effective mechanisms for regulation. The objective of this research is to compare the performance of water quality trading regimes regimes relying on technology-based and performance-based exchange, respectively.

In the case of the Clean Air Act Amendments of 1990, the regulatory agency shifted its focus toward performance-based regulation. While some authors (Burtraw, 1996, for example) argue that shifting to performance based standards reduces the costs of complying with environmental regulations, the nonpoint source trading literature has focused on technology-based management (Letson, 1992; Malik, *et al*, 1993; Malik, *et al*, 1994). Two problems with water quality trading markets based on technology rather than performance standards must be addressed. First, uncertainty surrounding the effectiveness of abatement technologies for reducing nonpoint source pollution is great. Malik *et al.* (1993) have shown that if this uncertainty is too large, it can eliminate the incentive for point sources to enter into trading agreements. Second, the trading ratio does not provide incentives to deal with asymmetric information problems. It is therefore important to assess how the trading ratio responds to problems of moral hazard and adverse selection, and the resulting effect on market efficiency.

The central hypothesis of the dissertation is that adopting performance standards rather than technology standards in point-nonpoint trading programs will provide for increased efficiency within the water quality trading market. To assess this hypothesis, I develop a theoretical economic model of a collective performance-based trading water quality trading market. A collective performance-based system would be expected to

transfer uncertainty from the regulator and the point sources to the nonpoint sources. This should improve efficiency, because the nonpoint sources are best suited to handle this uncertainty as they have more knowledge of the types of practices and technologies that could be used to abate pollution. Also, performance-based systems are incentive based, and therefore encourage the adoption of cost minimizing technologies.
The main innovation of this hypothesis is that it addresses the use of collective performance standards in water quality trading markets.

An enforceable baseline would be a desirable feature of any water quality trading market. However, the current differential treatment of nonpoint sources under the Clean Water Act prohibits compulsory nonpoint source abatement requirements through either abatement technology adoption, or individual emissions standards. To gain a better understanding of the legitimacy of such a ban, and whether an enforceable baseline is a feasible option, it is important to examine whether there is any normative basis for the nonpoint source exemption.

This dissertation is an examination of the difficulties encountered in extending tradable permit markets to include nonpoint sources. The inability to monitor individual nonpoint source emissions levels leads to two questions. The first is if nonpoint sources can be fully incorporated into the market, or if they must be brought in on a voluntary participation basis. The second is whether it is more efficient to accommodate trading between point and nonpoint sources through a technology-based or collective-performance based approach.

In Chapter 2, the technology-based approach to water quality trading is examined. In particular, the regulator's use of a trading ratio to deal with the uncertainty between abatement technology adoption and actual individual abatement is explored in terms of its efficiency and effectiveness. A safety-rule model of the trading ratio is developed which allows for a detailed discussion of both mean and variance effects in the presence of symmetric and asymmetric uncertainty.

In Chapter 3, the alternative of collective performance-based trading is introduced. Any contracting arrangement based on collective performance must overcome two asymmetric information problems. The first is adverse selection, where the point source cannot identify the low-cost nonpoint sources, and does not have enough information to ensure that both incentive compatibility and participation constraints will be met. Using mechanism design, a two-stage contract that pairs a team entry auction with an "all-or-nothing" team contract is developed, and its efficiency is discussed.

Chapter 4 explores whether there is any morally relevant distinction between point and nonpoint sources of pollution that justifies the exemption of nonpoint sources from the polluter-pays-principle. The distinction in terms of differential observability is explored under both technology-based and performance-based regulatory approaches. The role of collective penalties and tradeoffs between imposing costs on different classes of undeserving individuals are examined.

CHAPTER 2

TECHNOLOGY-BASED TRADING APPROACH

This chapter provides a general examination of the efficiency of technology-based water quality trading (WQT) markets. All current and proposed WQT markets rely on a technology-based approach to trading. Observable units of abatement technology adoption, such as filter strips or tillage practices, serve as a proxy for the unobservable level of abatement performed by individual nonpoint sources.

The introduction of a proxy for individual nonpoint source abatement allows the regulator to observe inputs in the abatement process, estimate reductions in individual nonpoint source pollution emissions, and provide a basis for enforcing noncompliance in terms of failure to adopt. However, the actual performance of nonpoint source abatement technology can vary depending upon a wide range of factors, including but not limited to, heterogeneous site characteristics.

To overcome this problem, technology-based WQT programs have introduced the concept of trading ratios. A trading ratio sets the rate at which point source emissions can be exchanged for expected nonpoint source emissions. When trading ratios are set at a value greater than one, more than one unit of expected nonpoint source abatement must

be produced to earn a single credit for the point source.[3] The use of trading ratios to deal with market uncertainty has consequences in terms of both economic efficiency, and environmental protection. For this reason, this chapter focuses primarily on the effects of the trading ratio on WQT market operation.

I begin by developing a model of the trading ratio as a safety rule, which differs from the traditional analysis of the trading ratio in the literature. This formulation allows for the direct examination of changes in both the mean and variance of total emissions, and the corresponding impact upon the optimal trading ratio.

If the adoption of nonpoint source abatement technology captures, albeit imperfectly, all of the relevant pollution control actions of the nonpoint source, trading based on technology adoption is tantamount to trading nonpoint source emissions. In this case, efficiency and effectiveness require only the consideration of how to set the optimal trading ratio to deal with the variability of emissions and abatement technology performance. This has been the focus of the economic literature on WQT programs to date (Taff and Senjem, Malik, et al, 1993; Horan, 2001). This approach treats the distinction between point and nonpoint source pollution as one of differential uncertainty. Point source emissions and technology performance are considered deterministic, while nonpoint source emissions and technology performance are considered stochastic.

Beginning with this assumption of symmetric information, I show that the safety rule version of the trading ratio is equivalent to the optimal trading ratio as developed in the economic literature. I also address recent arguments that economic efficiency

[3] For a trading ratio less than one, less than one unit of expected nonpoint source abatement must be produced to earn a single credit for the point source.

requires the optimal trading ratio always be set at less than one, which is contrary to what is seen in practice. Using the safety rule analysis, I provide two practical reasons why the optimal trading ratio may be set at greater than one, even in a symmetric information setting.

In most cases, however, both observable and unobservable actions affect the emissions and performance of nonpoint source abatement technology or practices. For example, some aspects of changes in tillage practices are easily observed, while other important dimensions are unobservable. Thus, a more useful formulation of the point-nonpoint distinction can be made in terms of asymmetric information. The technology-based approach as discussed in the literature, and adopted in practice, does not explicitly account for this distinction.

The trading ratio does not provide incentives for individual nonpoint sources to perform the socially desired, but privately costly, unobserved actions that improve abatement performance. The regulator's adjustment to this problem may be to set high, and potentially prohibitive, trading ratios, which can greatly hamper market efficiency. To examine this hypothesis, I conclude with an examination of the trading ratio in the presence of asymmetric information. The effects of hidden types and hidden actions are discussed in terms of their impact on WQT market efficiency and effectiveness. The failure to account for asymmetric information could undermine the usefulness of the current reliance on technology-based approaches.

2.1 A Model of Technology-based Trading Under Symmetric Information

Malik, et al, propose the following model of point and nonpoint source emissions: Consider a watershed where total emissions of uniformly mixing pollutant are drawn from two sources. One is a point source, and the other is a nonpoint source, $e_T = e_P + e_N$. The nonpoint source is a farmer with L acres of land in production. The emissions of the nonpoint source differ from those of the point source in two key ways. First, nonpoint source emissions are dispersed over all acres in production, and the emissions from any particular acre of land cannot be observed directly. Under the technology-based trading approach, per acre emissions are approximated based on the type of abatement technology being used. Second, the emissions of nonpoint sources are greatly influenced by weather factors. Therefore, aggregate nonpoint source emissions can be represented as $e_N = (L - L_1)g_o(\omega) + L_1 g_1(\omega, \varepsilon)$, where L_1 denotes the acres of land on which a particular abatement technology has been adopted. Per acre nonpoint source emissions without abatement technology is represented by $g_o(\omega)$, and per acre nonpoint source emissions with new abatement technology are $g_1(\omega, \varepsilon)$. The random variable ω represents stochastic weather factors that influence nonpoint source emissions from all acres of farmland. The random variable ε represents the uncertainty associated with the performance of the newly adopted abatement technology in reducing per acre emissions. It is assumed that both random variables are normally distributed, with mean zero and finite variance. Therefore, $E[g_o(\omega)] = g$ and $E[g_1(\omega, \varepsilon)] = g_1$. By definition, the

expected per acre emissions under the existing technology are greater than the expected per acre emissions under the abatement technology, i.e., $E[g_o(\omega)] > E[g_1(\omega,\varepsilon)]$, but there is no *a priori* restriction on the relationship between the variance of per acre nonpoint source emissions under both technologies, i.e., $Var[g_o] \overset{\leq}{\underset{\geq}{}} Var[g_1]$ [4]. The adoption of abatement technology reduces per acre nonpoint source emissions on average, but can increase or decrease the variability of nonpoint source emissions.

A water quality trading market allows the point source to offset an increase in its own emissions by purchasing reductions in nonpoint source emissions. In a technology-based approach, offsets are based on expected nonpoint emissions reductions from the adoption of a specified abatement technology. The point source is the only regulatory controlled source of emissions. The pre-trading permitted level of point source emissions is denoted as e_P. The assumption that point source emissions are deterministic, coupled with an assumption of a pre-existing optimal enforcement mechanism, implies that the point source will not violate, e_P.[5] The point source will increase emissions beyond e_P if abatement technology is adopted on enough acres so that expected nonpoint source emissions are reduced by an equal amount and at less cost. Within the water quality trading market, total allowable emissions can be allocated between the sources in various combinations, as in equation 2.1:

[4] This assumption also leaves open the correlation of the random variables.
[5] The assumption of deterministic point source emissions is a legacy of the models that have based the point-nonpoint source distinction on differential uncertainty. The effect of relaxing this assumption is examined later in the chapter.

$$e_T = e_P + L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)]) + (L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon) \quad (2.1)$$

Without trading, there is no adoption of nonpoint source abatement technology. Therefore, $L_1 = 0$, and total emissions are the sum of the permitted level of point source emissions plus the unrestricted nonpoint source emissions from the *L* acres of land in agricultural production $e_T = e_P + Lg_o(\omega)$. With trading, the point source emissions are increased by the expected reduction in nonpoint source emissions from the purchase of abatement technology on L_1 acres of farmland, $L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)])$. This increase is offset by actual nonpoint source reductions, $(L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon)$.

Due to the distributional assumptions regarding the means of ω and ε, trading has no effect on total expected emissions, as the increase in point source emissions equals the expected reduction in nonpoint source emissions:

$E[e_T] = E[e_P + L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)]) + (L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon)] = e_P + LE[g_o(\omega)]$.

However, trading does have an impact on the variance of total emissions, which, in this setting, is identical to the variance of nonpoint source abatement:

$Var[e_N] = (L - L_1)^2 Var(\omega) + L_1^2 Var(\omega)Var(\varepsilon) + Cov[\omega,\varepsilon]$. As mentioned previously, the variance of nonpoint source abatement increasing or decreasing with the adoption of new technologies is *a priori* ambigious.

Using a model of abatement cost minimization subject to expected damages being less than an exogenously set standard, the trading ratio is derived from the following efficiency condition:

$$\frac{\dfrac{\partial C_P(e_P)}{\partial e_P}}{\dfrac{\partial C_N(e_N)}{\partial e_N}} = \frac{\dfrac{\partial E[D(e_P,e_N)]}{\partial e_P}}{\dfrac{\partial E[D(e_P,e_N)]}{\partial e_N} - \lambda\left(\dfrac{Cov\left[\dfrac{\partial D(e_P,e_N)}{\partial e_N},(g_1-g_o)\right]}{E[g_o]-E[g_1]}\right)} \quad \text{(Malik, et al).}$$

When both point source and nonpoint source emissions are deterministic, this equation reduces to the typical efficiency condition of the ratio of marginal abatement costs equaling the ratio of marginal damages. With stochastic nonpoint source emissions, the marginal damages from nonpoint source emissions include a "marginal damage premium" that determines the optimal change in the trading ratio (Malik, et al.). Given a convex environmental damage function, if increased use of nonpoint source abatement increases (decreases) the variance of nonpoint source emissions, the trading ratio must also increase (decrease). This formulation of the trading ratio is useful in terms of the empirical derivation of the trading ratio. However, a more holistic approach to the optimal regulation of stochastic emissions provides a more illustrative basis for the optimal trading ratio under symmetric and asymmetric information.

2.2 The Safety Rule Approach

The safety rule approach has been shown to be an effective method for the control of stochastic environmentally damaging emissions (Beavis and Walker; Lichtenberg and

Zilberman; Lichtenberg, Zilberman and Bogens). When dealing with stochastic emissions, the safety rule approach allows the regulator to address the variability of emissions while regulating expected emissions. The safety rule corresponds to a disaster avoidance approach to decision-making. The likelihood of any deviations from the regulatory standard is constrained to occur within an acceptable level.

The safety rule approach bases regulatory decision-making on two parameters: maximum allowable risk, i.e., the expected total emissions standard, $E\left[e_T^*\right]$, and a margin of safety, α. This provides the practical advantage of utilizing decision parameters that are relatively easy for regulators to grasp conceptually, and to determine intuitively, as they tend to be numbers around which public debate centers (Lichtenberg and Zilberman).

A regulator who wishes to constrain violations of the regulatory standard, $E\left[e_T^*\right]$, to occur with some probability α, i.e., $\Pr\{e_T \le E\left[e_T^*\right]\} = \alpha$, could use the following safety rule to set the probabilistic standard, S_T, as: $S_T \ge E[e_T^*] + Z\sqrt{Var(e_T)}$. The letter Z denotes the standard normal variable corresponding to the given probability level α.[6] The probabilistic safety rule can be broken into two parts. The first part ($E\left[e_T^*\right]$) is the mean of total emissions, and the second part ($Z\sqrt{Var(e_T)}$) can be called the safety margin. Since all uncertainty in the model is assumed to restricted to nonpoint source emissions, the safety margin can be rewritten as, $Z\sqrt{Var(e_N)}$. The safety rule allows

[6] Z is derived from the following condition: $\alpha = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}u^2\right) du$

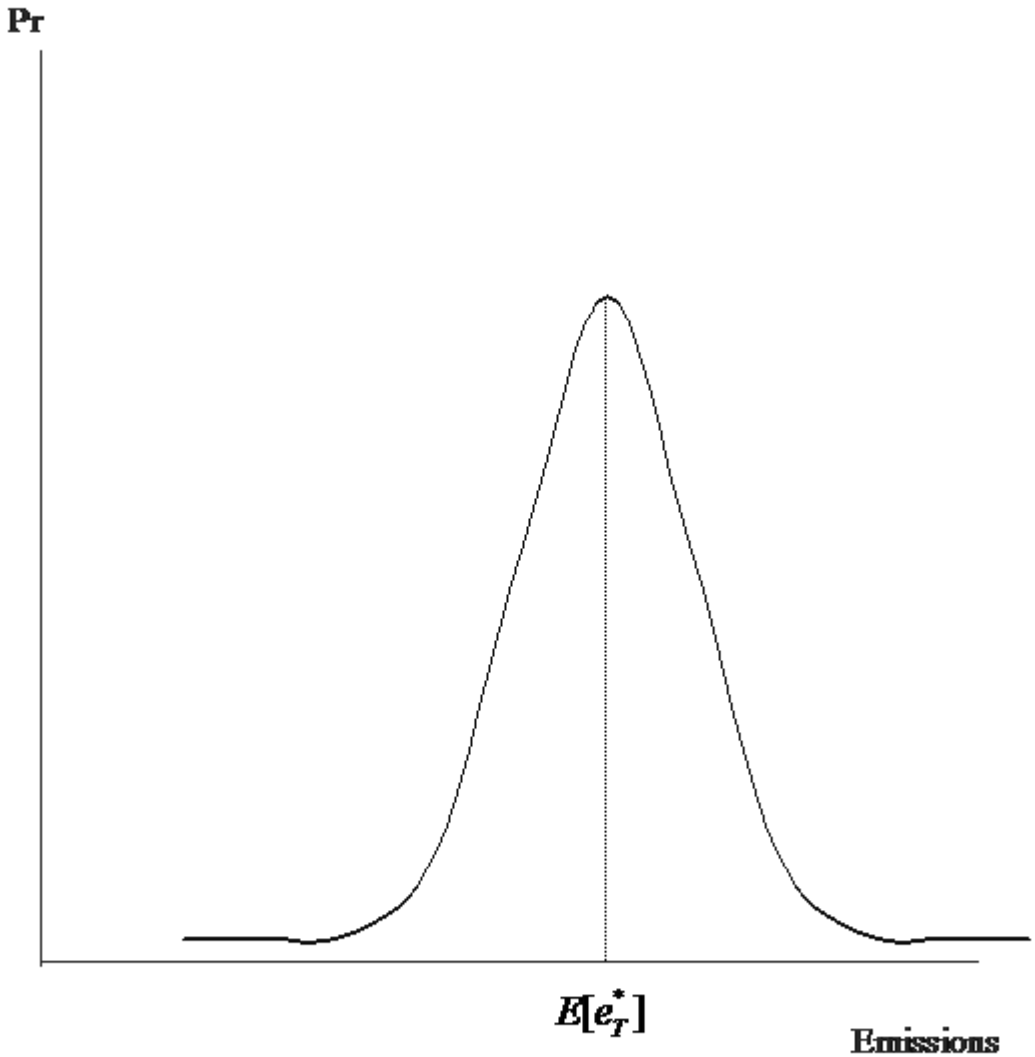

Figure 2.1: Risk Neutral Safety Rule for Stochastic Emissions

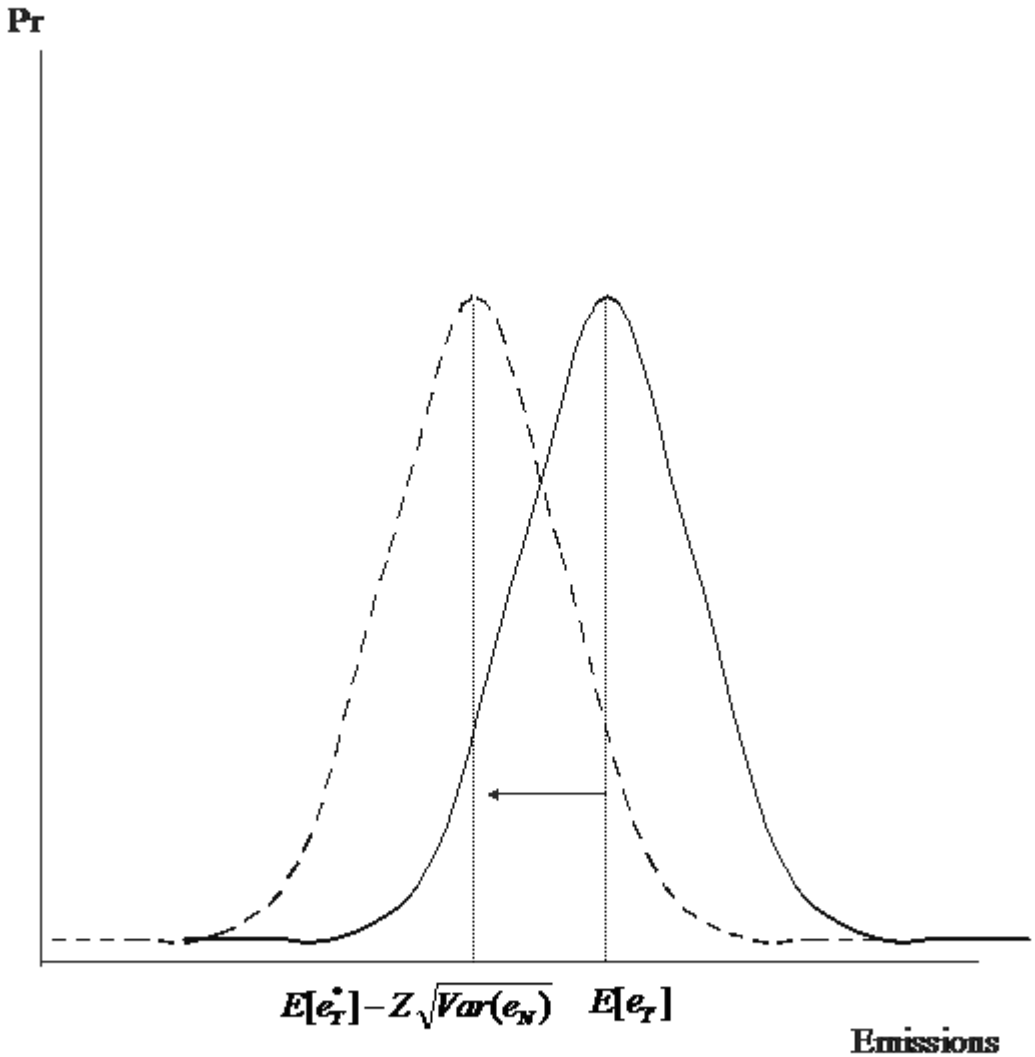

Figure 2.2: Risk Averse Safety Rule for Stochastic Emissions

trading based on expected emissions by adjusting the expected level of permitted emissions to account for the corresponding change in the variance of total emissions. This prevents violations of the regulatory standard from occurring with more than an acceptable level of frequency.

This formulation of the decision process as a combination of mean risk and uncertainty allows for some interpretation of the regulators' aversion to uncertainty, which is the counterpart of the traditional notion of risk aversion in this context (Lichtenberg and Zilberman). A risk neutral regulator will want to ensure that, on average, the emissions standard is met. Since it is assumed that the uncertainty is distributed symmetrically around the mean, the risk neutral regulator does not have to adjust the expected emissions standard. In this case, Z=0 and $\alpha = 0.5$. Trading will occur based simply on expected loadings.

The risk averse regulator will want to ensure the target is met with greater frequency than the risk neutral regulator, thus setting α>0.5. This increase in α, results in Z<0, and a safety margin that decreases the level of allowable expected emissions. In Figure 2.2, the reduction of the mean level of total allowable emissions to $E[e_T^*] + Z\sqrt{Var(e_N)}$ shifts the entire distribution of total emissions to the left. The desired level of expected emissions, $E[e_T^*]$ is achieved with a higher level of probability (i.e., the area under the shifted (dashed) probability distribution function to the left of $E[e_T^*]$).

In developing a model of an economically the optimal trading ratio, it is assumed that the regulator seeks to maximize social welfare, and is only derivatively interested in

emissions.[7] The regulator is only concerned with how emissions affect environmental damages. A risk neutral regulator would require that potential losses (the change in environmental damages when emissions exceed the expected level) equal the potential gains (the change in environmental damages when emissions fall short of the expected level). The safety rule is now transformed to $\Pr\{D(e_T^*) \leq D(E[e_T^*])\} = \alpha$, and $D_T \geq D\left(E[e_T^*] \pm Z\sqrt{Var(e_N)}\right)$. The standard is now based on adjusted allowable damages, which we can write for convenience in terms of emissions as: $S_D = E[e_T^*] \pm Z_D\sqrt{Var(e_N)}$, where the subscript D is meant to remind the reader that the probabilistic standard and the sign and size of the safety margin are determined with respect to the environmental damage function.

In the special case of a linear environmental damage function, regulating environmental damages is the same as regulating emissions. Figure 2.3[8] depicts the constant marginal environmental damage curve for a linear environmental damage function. Since the marginal environmental damage curve is constant, damages are a monotonic transformation of emissions. and the regulator makes the same control decision in both cases. The risk neutral regulator would simply require mean total emissions to equal the mean allowable risk standard. The increase in damages when

[7] The choice of a welfare maximizing regulator translates the safety rule into an expected welfare rule. However, agencies charged with implementation of public policy may have alternative or at least competing goals. The utilization of a different goal would affect the choice of α and Z, but would not change how the safety rule is applied.

[8] The figures in the remaining portion of this chapter are used to provide a simplified visual description of the safety margin concept. For the sake of clarity, the continuous normal distribution of emissions, as described in the text, is represented by a discrete binary distribution. In the figures, e_T^- and e_T^+ represent the equally probable deviation from the mean level of emissions. The variance of the distribution is represented by the spread between the end points.

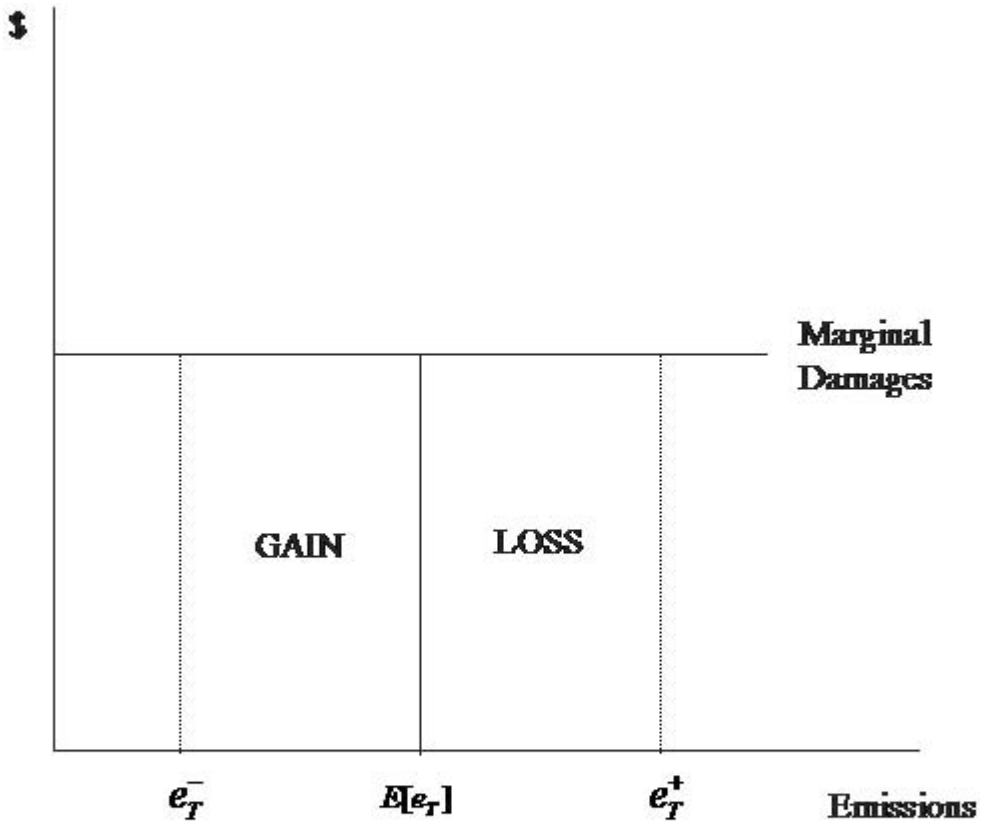

Figure 2.3: Stochastic Damages Loss and Gain – Linear Environmental Damage Curve

realized emissions are greater than the expected level will always equal the decrease in damages when realized emissions are less than the expected level. Thus, the risk neutral regulator sets α=0.5, which, in the case of constant marginal damages, results in Z=0, because the symmetry of emissions uncertainty is maintained in the symmetry of environmental damages. Trading is allowed based on expected emissions, without a safety margin, because of the symmetry of potential losses and gains.

An environmental damage function that is convex in emissions will have an upward sloping marginal environmental damage curve. Again, a risk-neutral regulator will choose to set α=0.5, i.e., expected losses and expected gains in environmental damages occur with equal probability. However, the curvature of the environmental damage function will require Z to be set differently than in the emissions problem.

The upward sloping marginal environmental damage curve implies that symmetric uncertainty in emissions will result in expected losses being greater than expected gains (Figure 2.4). The expected loss when realized emissions are greater than the mean are represented on the graph as the area under the marginal damage curve between the mean, $E[e_T^*]$, and the upperbound e_T^+. Likewise, the expected loss is represented by the area between the mean and the lower bound, e_T^-. Although the deviations from the mean are symmetric, the shape of the marginal damage curve creates differential impacts in terms of over- and under-production of emissions. In particular, expected losses are always greater than expected gains, and the larger the variance of total emissions the greater this disparity becomes. As a result, the risk neutral regulator will choose Z<0 and set a safety margin that decreases allowable total expected

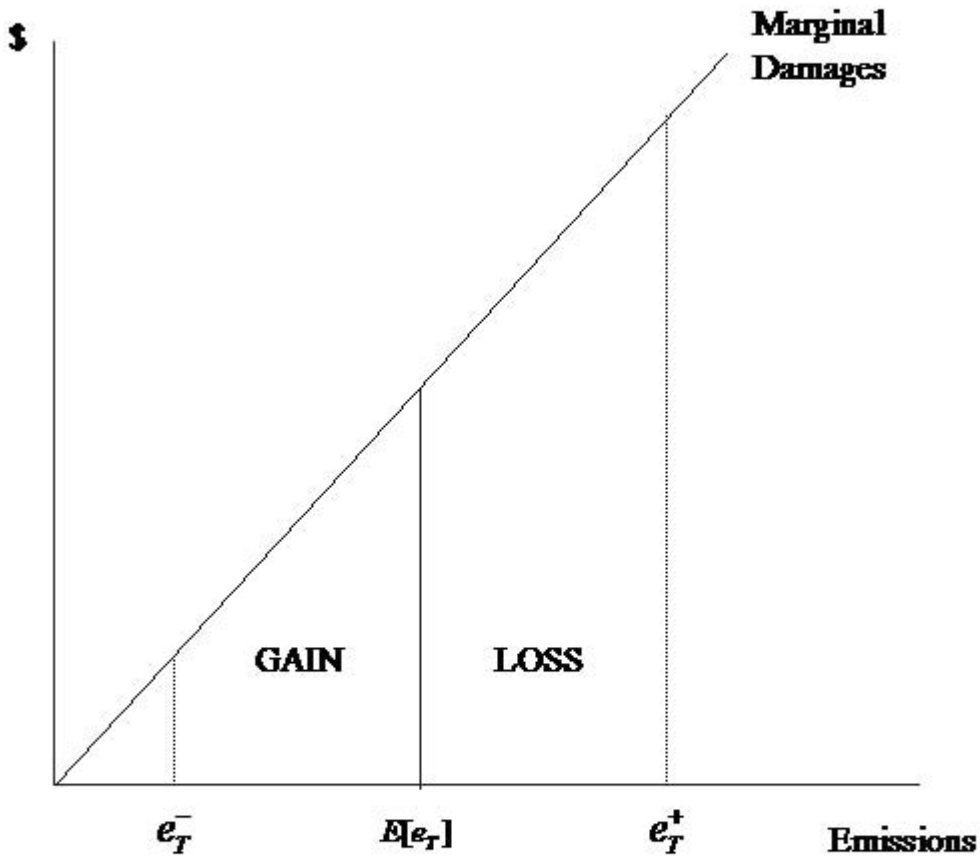

Figure 2.4: Stochastic Damages Loss and Gain – Convex Environmental Damage Curve

emissions, $S_D = E[e_T^*] - Z_D\sqrt{Var(e_N)}$ (Figure 2.5). The safety margin is shifting the entire distribution of emissions to the left. In this case Z_D must account for the impact of the stochastic emissions in terms of the environmental damage function. Given the variance of total emissions, Z_D is chosen to shift the distribution of emissions, such that the desired mean will be achieved with equal expected losses and gains. In Figure 2.5, the shift increases the expected gains (the area shaded by horizontal lines) and decreases the expected losses (the area shaded by vertical lines). At the adjusted, lower, emission standard, $E[e_T^*] - Z_D\sqrt{Var(e_N)}$, the desired level of emissions, $E[e_T^*]$, will be achieved with equal expectations of losses and gains. A concave environmental damage function will have a downward sloping marginal environmental damage curve. A downward sloping marginal environmental damage curve will result in a safety margin in the opposite direction. The regulator chooses Z>0 and sets a safety margin that increases allowable expected emissions $S_D = E[e_T^*] + Z_D\sqrt{Var(e_N)}$.[9]

A risk neutral regulator, facing uncertainty in total emissions, must choose both the sign of the safety margin, $\pm Z_D$, and the size of the safety margin. The sign of the safety margin is determined entirely by the curvature of the environmental damage function, and establishes whether the total emissions standard will be set at greater or less than the expected level. The size of the safety margin is decided based on the variance of total emissions, and any reallocation of emissions between the point and nonpoint sources

[9] The concave environmental damage function is included for completeness, but is not discussed further in the chapter, as it is unlikely to be a realistic representation of water pollution.

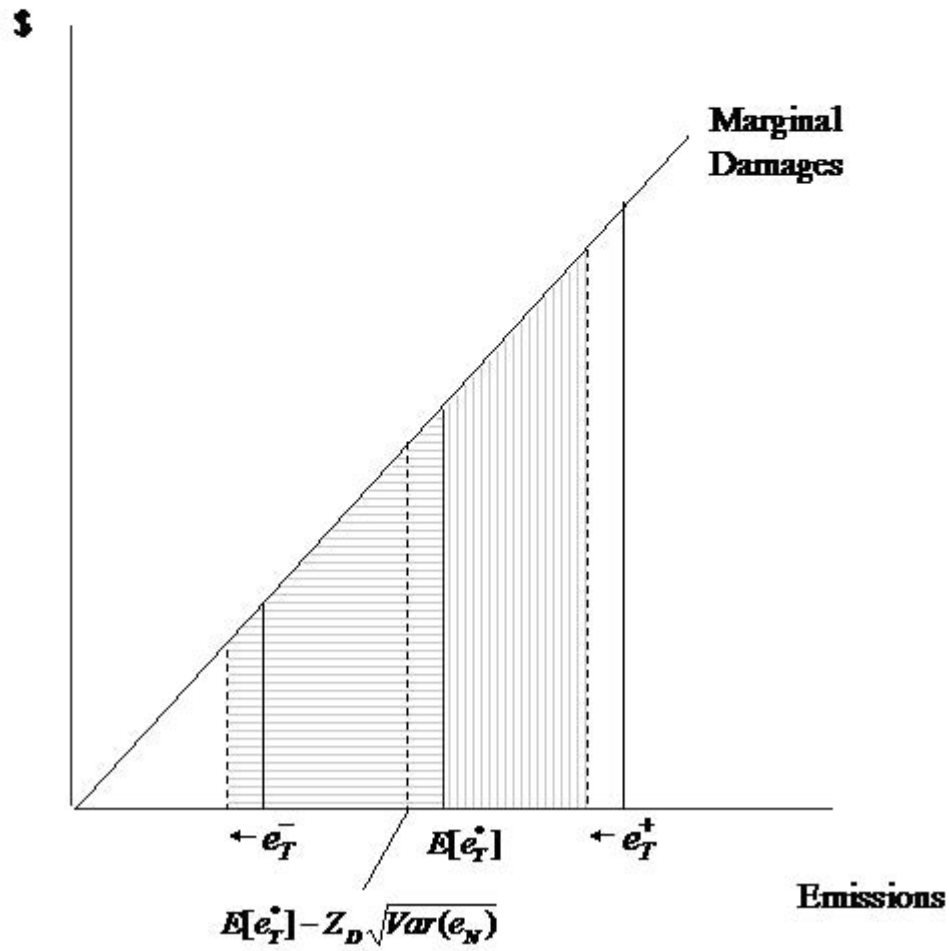

Figure 2.5: Safety Rule for Stochastic Damages – Convex Environmental Damage Curve

will require an adjustment. Unless otherwise stated, a convex environmental damage function is assumed throughout the remainder of the chapter.

Up to this point, the analysis has concentrated on developing the optimal safety margin to deal with stochastic emissions as they affect damages in a static sense. A risk neutral regulator facing a convex environmental damage function will always choose to reduce the expected ambient emissions standard to ensure that expected losses and expected gains are equal. We now turn to the question of how the size of the safety margin will change in response to changes in the variance of total emissions, which, as previously shown in Section 2.1, has been the focus of the trading ratio literature to date.

2.3 Linking the Safety Rule and the Trading Ratio

The relationship between the optimal trading ratio and the safety margin can be revealed using the following model, which assumes a convex environmental damage function ($\frac{\partial D(e_T)}{\partial e_T} < 0$) and a risk neutral regulator (α=0.5). From equation 2.1, total emissions prior to trading are $e_T^0 = e_P + E[Lg_o(\omega)]$, which will require the regulator to set the total allowable emissions standard as: $S_D^0 = e_P + E[Lg_o(\omega)] - Z_D\sqrt{Var(e_N^0)}$. The safety margin at the baseline, $-Z_D\sqrt{Var(e_T^0)}$, will be denoted as M^0. Since nonpoint sources are not directly regulated, it is assumed that the optimal safety margin at the baseline is achieved through the use of nonpoint source cost-share subsidy programs.

Trading reallocates emissions between point and nonpoint sources, and the allowable level of total emissions that maintains the baseline level of protection is denoted as:

$S_D^1 = E\left[L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)]) + (L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon)\right] + e_P - Z_D\sqrt{Var(e_N^1)}$. The new safety margin, corresponding with the reallocation of emissions from trading $Z\sqrt{Var(e_N^1)}$ is denoted as M^1. Finally, the change in the safety margin due to trading can be denoted as $\Delta M = M^1 - M^0$, and the following conditions hold:

If $Var(e_T^1) > Var(e_T^0)$ then $\Delta M > 0$

If $Var(e_T^1) < Var(e_T^0)$ then $\Delta M < 0$.

To determine the consistency between the safety rule approach to the trading rule approach, the following equation compares the safety margin as is determined by each. In the right hand side of the equation, T represents the trading ratio:

$$e_P + L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)]) + (L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon) - (M^0 + \Delta M) =$$
$$e_P + \frac{L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)])}{T} + (L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon) - M^0$$

This equality condition can be simplified to $\Delta M = L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)])(1 - \frac{1}{T})$,

which implies the following relationships between the safety rule and the trading ratio:

If $Var(e_T^1) > Var(e_T^0)$ then $\Delta M > 0$ and T>1

If $Var(e_T^1) < Var(e_T^0)$ then $\Delta M < 0$ and T<1.

Therefore, the safety rule provides results consistent with the trading ratio as developed in the literature (Taff and Senjem; Shortle, 1990; Malik, et al, 1993; Horan,

2001). The purpose of the trading ratio is to ensure that offsetting nonpoint source emissions with point source emissions does not violate the baseline safety margin. If offsets make total emissions more variable, then the safety margin must be increased, and the trading ratio is greater than one. If offsets make total emissions less variable, then the safety margin can be decreased, and the trading ratio is less than one.

The trading ratio is a marginal concept. It informs us about how to respond to changes in the variance of total emissions (i.e., whether the safety margin must be increased or decreased). It does not, however, provide any insight into the overall safety margin requirement, which is determined with reference to the shape of the environmental damage function. The failure to explicitly note this has led to some confusion when interpreting the optimal size of the trading ratio in practice.

2.4 Should the Optimal Trading Ratio Always be Less Than One?

In practice, all technology-based WQT programs use trading ratios that are greater than one (Woodward, 2001). In contrast to this, Horan (2001) argues that, in fact, the economically optimal trading ratio should always be set less than one, and that the trading ratios seen in practice are at best a result of confusion regarding the nature of environmental risks, and at worst, as overt concessions to the political influences of environmental interest groups. The costs of such confusion or concession are increased levels of social risks from water quality degradation due to sub-optimal reductions in nonpoint source pollution.

In the following section, I make the same simplifying assumptions regarding the nature of nonpoint source emissions as in Horan (2001). First, only weather uncertainty is considered in the model. Second, the mean and the variance of nonpoint source emissions are assumed to be positively correlated. A reduction in the mean level of nonpoint source emissions unambiguously reduces the variance of nonpoint source emissions. Finally, weather affects the emissions from acreage with newly adopted abatement technology in the same way as it affects emissions from acreage with existing technology. Under these assumptions, Horan (2001) argues that the economically efficient trading ratio must always be set less than one. I will show that this result holds only under very limiting assumptions regarding uncertainty and informational symmetry.

Let total emissions be represented as $e_T^1 = (E[e_n] - E[\phi])(1+\omega) + e_p + E[\phi]$, where ϕ is the contracted reduction in nonpoint source emissions and ω is the random variable representing weather effects that is distributed normally with mean zero and a finite variance. As in the previous section, the regulator will determine the appropriate baseline safety margin: $S_D^0 = E[e_N] + e_P - M^0$, where $M^0 = Z_D(E[e_n^2])\sqrt{Var(\omega)}$. When trading is introduced, the regulator adjusts the safety margin accordingly. With trading, the variance of total emissions changes, and the safety margin becomes: $M^1 = Z_D(E[e_N{}^2 - \phi^2])\sqrt{Var(\omega)}$. Offsetting stochastic nonpoint source emissions with certain point source emissions now unambiguously reduces the variance of total emissions, and justifies a reduction in the size of the safety margin, such that $\Delta M = M^1 - M^0 = Z(E[e_N - \phi])^2\sqrt{Var(\omega)} - Z(E[e_n^2])\sqrt{Var(\omega)} = -Z\phi^2\sqrt{Var(\omega)}$.

Therefore, a trading ratio less than one is optimal. Figure 2.6 illustrates the effects of a change in the variance of total emissions, due to trading, on the optimal safety margin. The marginal cost curve without trading is represented by MC^0. This can be thought of as the point source's marginal cost of emissions in the absence of any nonpoint source abatement. The marginal costs curve with trading is denoted MC^1, and is the marginal cost of emissions given that the lowest cost mix of point and nonpoint source abatement are implemented. For a WQT market to be feasible, the condition $MC^1 < MC^0$ must hold over some range of abatement. The assumption of trading leading to reduced variance in total emissions is depicted in terms of the upper and lower bound realizations. Thus, $e_T^{-^0}$ and $e_T^{+^0}$ have a greater spread than $e_T^{-^1}$ and $e_T^{+^1}$. The larger variance in total emissions implies that the expected loss/gain disparity is greater without trading than with trading. Therefore, with trading, a smaller safety margin is required to meet the standard, $M^1 < M^0$. This in turn, means that the optimal trading ratio should be less than one, as less than one unit of expected nonpoint source emission can be exchanged for a one unit increase in point source emissions, and the optimal safety margin will still be achieved.

This analysis shows that the safety rule approach confirms Horan's conclusion that when trading decreases the variance of total emissions, the optimal trading ratio should be set less than one. However, this result is dependent upon some limiting assumptions that, when relaxed to fit the conditions that are likely to exist in practice, reveal a different set of results. While it is conceded that political considerations are likely to have some bearing on setting the trading ratio in WQT markets in practice, it

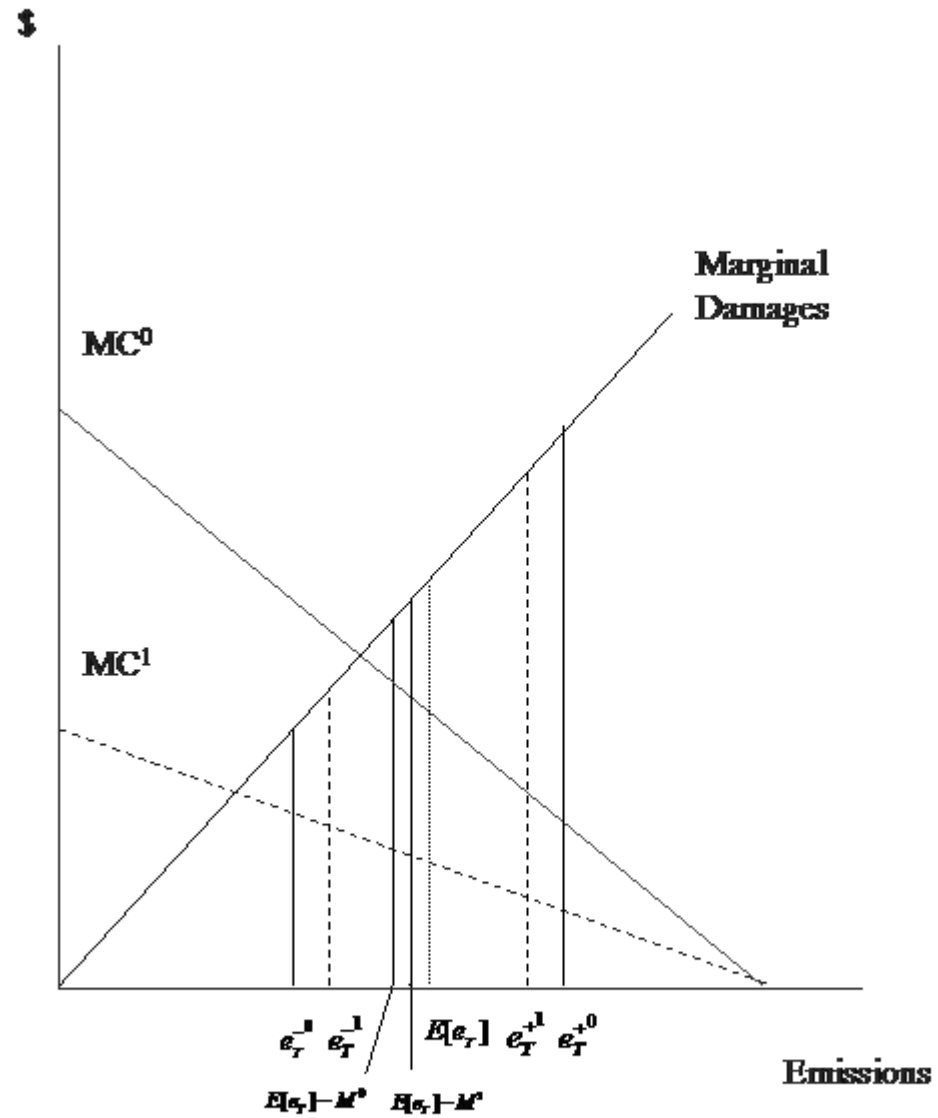

Figure 2.6: Optimal Trading Ratio When Trading Decreases the Variance of Total Emissions

does not follow that this is the only reason for the presence of trading ratios greater than one in practice. In fact, it is arguable that the optimal trading should always be greater than one in practice.

2.5 Two Reasons the Optimal Trading Ratio May Be Greater Than One

The results of the previous economic literature that the trading ratio must be less than one when the variance of total emissions is decreasing, and vice versa, are dependent upon at least two strong assumptions. The first is an implicit assumption regarding full regulatory compliance prior to trading. The second is the explicit assumption that point source emissions are deterministic. In this section, these assumptions are relaxed to be more consistent with the actual conditions present in WQT markets. It is shown that, even under an assumption of decreasing variance of total emissions, the optimal trading ratio is likely to be greater than one in practice.

In all of the previous literature, it has been assumed that trading is being introduced in a situation of compliance, that is that the ambient water quality conditions are being met prior to trade. Although this is a reasonable assumption for an economic model, it is unrealistic in practice. As mentioned in the introduction, the use of WQT programs is a result of the US EPA's desire to find innovative solutions for the large number of watersheds that are not currently meeting their water quality designations. Therefore, the watersheds in which these programs are being established have not achieved baseline standards from the outset.

As in the previous section, assume that the regulator sets the pre-trading allowable emissions as $S_D^0 = E[e_N] + e_P - M^0$, where $M^0 = Z_D(E[e_n^2])\sqrt{Var(\omega)}$. However, now consider the case in which the regulator is unable to reduce nonpoint source emissions through cost-share programs, $M^0 = 0$ [10]. As a consequence, the watershed fails to meet its ambient water quality standard prior to trading occurring. When trading is introduced, the regulator sets the safety margin as shown above, at $M^1 = Z(E[e_N - \phi])^2\sqrt{Var(\omega)}$. The pre- and post-trading safety margins are identical to the example used before to show the optimal trading ratio as less than one. However, in this case of noncompliance the change in the safety margin is not from M^0 to M^1, as the baseline safety margin was not achieved. Instead the change in the safety margin is from $E[e_N] + e_P$ to $E[e_N] + e_P - M^0$. The existing noncompliance requires the safety margin to be increased, even though variance is decreasing, and thus the trading ratio must be set greater than one.

Therefore, even though the variance of total emissions is decreasing, the optimal trading ratio is set greater than one. In the end, both the compliant case, with the trading ratio greater than one, and the noncompliant case, with the trading ratio greater than one, achieve the same optimal safety margin, $M^1 = Z(E[e_N - \phi])^2\sqrt{Var(\omega)}$. The only difference is that in the compliant baseline case, the move to the post-trading safety margin represents a decrease from the actual, existing safety margin, and in the noncompliant case it represents an increase from the actual, existing safety margin.

[10] This is the case of total noncompliance in terms of the safety margin. In cases, where noncompliance is only partial, $M^0 < M < 0$, the optimal trading ratio can be greater than or less than one, depending upon the impact trading has on total variance. Even in cases where the trading ratio will still be less than one, it will be much smaller than the optimal trading ratio under the compliance assumption.

A second reason for believing that the optimal trading ratio could be greater than one in practice relates to the emissions of point sources. The assumption that point source emissions are deterministic does not correspond to actual conditions. For example, wastewater plants that treat combined sewage overflows can have emissions greatly affected by high precipitation events, and all abatement technology is susceptible to unexpected breakdown. In addition, although individual point source violations can be detected *ex post* in theory, in reality very few noncompliant firms are detected each year. Prohibitively high monitoring costs preclude the ability to detect all non-compliant firms. Thus, in practice, point source emissions are both ex ante and ex post uncertain.

Relaxing the deterministic point source assumption has a significant impact on the optimal trading ratio, as the variance of total emissions is now determined by the combined variances of both the point and nonpoint sources, i.e., $\pm Z\sqrt{Var(e_T)}$. Thus, even if adopting nonpoint source abatement technology reduces the nonpoint source emissions variance, this must be weighed against the effect of decreased point source abatement on the variance of point source emissions. Not only are the mean-variance correlations for point and nonpoint source emissions unknown *a priori*, but their combined effects and relative magnitudes are also empirical facts that need to be determined prior to resolving whether the optimal trading ratio should be greater than or equal to one. Therefore, one does not need to appeal to political economy explanations for the presence of trading ratios greater than one in practice. The trading ratio may be greater than one, even if increased nonpoint source abatement results reduces the variance of nonpoint source emissions, if the pre-trading level of ambient water quality is

non-compliant, or if the uncertainty regarding point source emissions is acknowledged in the model.

2.6 A Model of Technology-based Trading Under Asymmetric Information

Differential uncertainty between point and nonpoint source pollution cannot serve as the foundational concept for nonpoint source regulation and the optimal trading ratio. The real issues in nonpoint source pollution, and the reasons why we see trading ratios greater than one in practice have to do in large part with problems of asymmetric information.

Nonpoint source pollution is diffuse, and therefore, individual contributions to overall emissions are difficult to observe. In the previous literature on water quality trading, it is assumed that this observation problem is symmetric. That is, the point sources, nonpoint sources, and the regulator all possess the same information regarding the emissions of individual nonpoint sources. Specifically, this would mean that they all know the distribution of per acre emissions, and the distribution of per acre reductions given adopted technology. However, in reality this is not the case. Nonpoint sources hold private information that is relevant to determining their individual contribution to overall emissions. This information includes, among other things, physical characteristics of their property (i.e., location, slope, soil type), and operational characteristics of their enterprise (nutrient application rates and timing, maintenance and management effort). The first category can be thought of as information known to the

nonpoint source prior to trading, i.e., hidden types. With minimal effort and expense much of this type of information can be obtained by the regulator. The second category can be thought of as actions that can be taken by the nonpoint source after trading has occurred, i.e., hidden actions. This information is much harder for the regulator to obtain, and because the technology-based trading approach does not provide incentives for the voluntary revelation of this private information, market efficiency is reduced.

In practice, many technology-based WQT programs have adopted the traditional approach to hidden type problems, by requiring some form of third party inspection of abatement technology. For example, many programs require trades to be approved by state or regional EPA offices prior to credits being issued. The State of Michigan Water Quality Trading Rules require that a licensed and independent third party certify the credits that should be earned from the adoption of each trade (US EPA, 2003). Alternatively, the use of GIS-based databases has been proposed as a lower-cost alternative to third party verification (WRI, 2000). These programs compile the relevant information regarding heterogeneous characteristics for all parcels in the watershed, and provide simulations of site-specific expected abatement performance from various technologies. However, no matter how careful the inspection or the simulation effort, some information will remain hidden. In addition, these methods for overcoming the hidden type problem add substantial transaction costs to the market, and will also reduce overall efficiency.

A more insidious problem in technology-based trading is that of hidden action. The technology-based trading approach has been favored because it provides an easily

observed proxy for individual emissions. How well the abatement technology approximates actual reductions plays a crucial role in market efficiency and effectiveness. The efficiency of the trading ratio developed earlier in this paper is dependent upon the random nature of both nonpoint source emissions, and abatement technology performance. When the nonpoint source can control any unobservable aspect of abatement technology performance, efficiency is threatened.

Almost all nonpoint source abatement technologies have the potential for abuse due to hidden action. When actual performance depends on actions other than technology adoption, the potential for Potemkin[11] abatement technologies exist. These are technologies that are perceived to be highly functional, but in actuality are underperforming. If the appearance of the technology is not an accurate measure, then it does not reflect actual performance and the technology-based trading market can be undermined.

For the analysis of hidden action, it is assumed that the environmental damage function is linear. Under the symmetric information assumption, the trading ratio responds to changes in the variance of total emissions only, which, given linear damages, will be set equal to one (see Figure 2.3). However, the impact of hidden action is realized at the mean rather than the variance of total emissions. Assuming a linear environmental

[11] Grigori Potemkin was Catherine the Great's governor to the Crimea. As legend has it, Potemkin constructed elaborate mock villages throughout the region, so as to create an illusion of prosperity when Catherine toured the countryside with foreign dignitaries.

damage function, clearly distinguishes the mean impact effect from the variance effect on the trading ratio, isolating the effect of hidden action.[12]

Under the symmetric information assumption, a safety margin is not required and the trading ratio is set equal to one, because observed variations in emissions and the performance of abatement technology are completely random and follow a symmetric distribution. Suppose that the nonpoint source can reduce abatement costs by taking (or choosing not to take) some hidden action. The outward appearance of the adopted technology is unaffected by the choice, but its effectiveness will be decreased, i.e., increased per acre emissions. Therefore, the per acre emissions from the adoption of new abatement technology can be rewritten as $g_1(\omega,\varepsilon)$ and $g_1^S(\omega,\varepsilon)$, where the superscript S denotes per acre emissions that result from the hidden actions of the nonpoint source. By definition, the expected level of per acre emissions is greater when the nonpoint source "shirks" by taking hidden actions, i.e., $E\left[g_1^S(\omega,\varepsilon)\right] > E\left[g_1(\omega,\varepsilon)\right]$.

Under this scenario, the optimal trading ratio derived under the symmetric information assumption will fail to protect the environmental standard. On average, increases in point source emissions will exceed nonpoint source reductions. Under the linear environmental damage function assumption, the allowable total emissions are simply: $S_D = e_P + LE\left[g_o(\omega)\right]$. Under the symmetric information assumption, any reallocation of emissions between the point and nonpoint sources meets the standard:

$$S = E\left[e_P + L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)]) + (L - L_1)g_o(\omega) + L_1 g_1(\omega,\varepsilon)\right] = e_P + LE[g_o(\omega)],$$

[12] The conclusions drawn under the linear damage function are directly applicable under either a convex or concave damage function assumption.

because, on average, the increase in point source emissions is offset by the corresponding decrease in expected nonpoint source emissions. However, the ability to take hidden actions will undermine this result:

$$\begin{aligned} S &= E\left[e_P + L_1(E[g_o(\omega)] - E[g_1(\omega,\varepsilon)]) + (L - L_1)g_o(\omega) + L_1 g_1^S(\omega,\varepsilon)\right] \\ &= e_P + LE[g_o(\omega)] + L_1 E\left[g_1^S(\omega,\varepsilon) - Eg_1(\omega,\varepsilon)\right]. \end{aligned}$$

Therefore, on average, the standard will not be met under the traditional trading ratio. To ensure compliance, a safety margin needs to be adopted (Figure 2.7). The sign of the safety margin is unambiguously negative, because shirking will only cause an increase in expected per acre emissions. The size of the safety margin, $L_1 E\left[g_1^S(\omega,\varepsilon) - Eg_1(\omega,\varepsilon)\right]$, is determined by the amount of shirking that can feasibly go unnoticed. In the best case scenario, the inability to provide an incentive to avoid shirking results in a loss of efficiency. In the worst case scenario, the degree of shirking is large enough to require a trading ratio that prohibits trading, and the market collapses.

The effects of hidden action impact the mean of expected total emissions, and unambiguously require the trading ratio to be set greater than one. In the case of a convex environmental damage function, the interaction of the mean and variance effects will determine the size of the optimal trading ratio. Although taking into account the previously discussed considerations, it is likely that the trading ratio will always be greater than one.

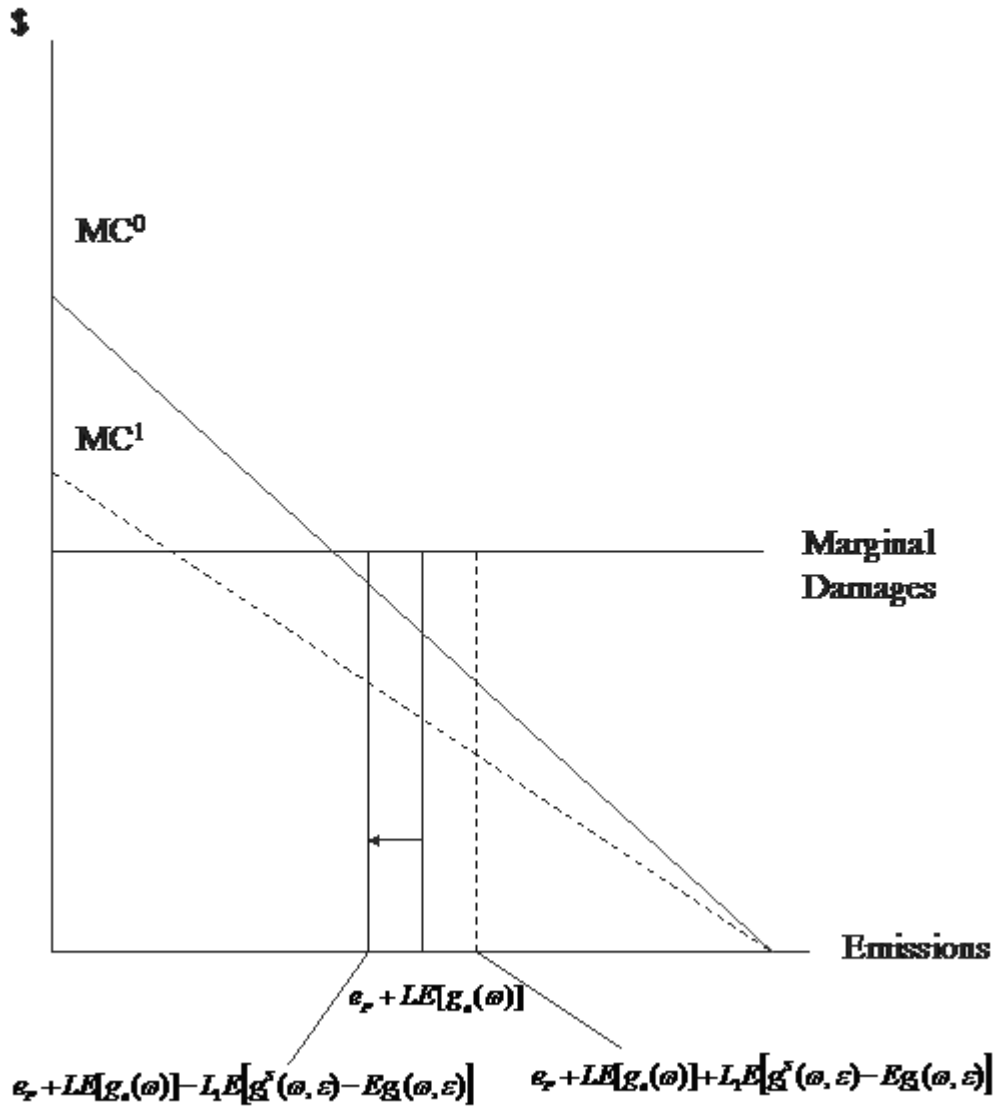

Figure 2.7: The Mean Effect of Hidden Action on the Optimal Trading Ratio

2.7 Conclusions

Difficulty observing individual levels of nonpoint source emissions poses a special challenge to WQT markets. When all parties involved (point source, nonpoint sources, regulator) have the same level of information regarding nonpoint source emissions and abatement, technology-based trading approaches can provide an efficient means for establishing trading. Trading ratios can effectively overcome the uncertainty associated with using a proxy for actual nonpoint source emissions. Under this assumption of symmetric information, the optimal trading ratio will respond to changes in the variance of total emissions that occur with reallocation of emissions among point and nonpoint sources.

In practice, various realities will affect the determination of the optimal trading ratio. Even when the variance of total emissions decreases with trading, contrary to the previous literature, the optimal trading ratio may be greater than one. This is particularly true when trading is being established in a noncompliant watershed.

The possibility of asymmetric information of hidden type or hidden action presents a serious challenge to the efficiency and effectiveness of technology-based trading. Hidden action allows individuals to shirk in the production of abatement through the given technology that is adopted. This impacts the trading ratio by requiring a shift in the expected level of total allowable emissions, requiring it to be set at a number greater than one. At a minimum, the presence of hidden type and action add additional costs to the WQT market reducing efficiency. More troublesome is the potential for hidden

action impacts to require trading ratios to be set so high as to remove all benefits of trade, causing the market to collapse.

Technology-based trading's reliance on the trading ratio is unsuccessful in the presence of asymmetric information. This is because it fails to provide incentives to align the concerns of the point and nonpoint sources. For example, making payments to nonpoint sources dependent upon their performance would reduce the incentive to cut costs by taking advantage of hidden action. This move to performance-based trading also has the additional benefit of increased efficiency. Under technology-based trading, the point source is able to achieve some measure of static efficiency, in the sense that it is able to capitalize on cost differences between its own abatement technology and that dictated to the nonpoint sources. However, dictating technology to be adopted precludes one of the most touted benefits of market-based instruments, and that is the dynamic efficiency gained from induced innovation. By focusing on performance as opposed to technology, the WQT market allows for incentives to overcome asymmetric information and holds the promise of increased efficiency through induced innovation. However, performance in the nonpoint source case is only observable at the collective level, which complicates its use as a basis for trade, which is the subject of the next chapter.

CHAPTER 3

COLLECTIVE PERFORMANCE-BASED TRADING APPROACH

The purpose of this chapter is to develop an efficient team contract that can be used as the basis of exchange within a collective performance-based water quality trading (WQT) market. A team contract will allow a point source to purchase abatement from multiple nonpoint sources whose actual abatement production is directly unverifiable. Individual contractual obligations are enforced through the verifiable performance of the entire team. A secondary goal is to design a practical team contract, which can be implemented for use with agricultural nonpoint sources. This requires a balancing of simplicity and transparency regarding contract details with the required incentive compatibility features. Therefore, the contract must be based on information readily available to agricultural producers, and it must correspond with some minimal criteria of fairness.

I begin the chapter with a discussion of the asymmetric information problems associated with team contracting: adverse selection and moral hazard. Nonpoint sources have different abilities in terms of the cost of reducing emissions. These abilities are privately known, but unobservable to others, which causes adverse selection problems for the point source who wishes to contract with the lowest cost group of nonpoint sources. Nonpoint sources are also better informed regarding their own unique abatement

production processes, and therefore hold private information regarding their actual production and contribution to collective levels of abatement. This creates a moral hazard problem, as the point source cannot observe if individual nonpoint sources are in breach of contract.

Next, a two stage contract is proposed that address both of these asymmetric information problems. An important distinction between the proposed contract and other team contracts in the literature is the use of a team entry auction to address adverse selection. This promises to overcome many of the criticisms that have been levied against the practical use of previous team contracting arrangements.

Finally, the efficiency of the team contract and team auction will be examined. The team contract must transmit the appropriate incentives to ensure that individual, and unverifiable abatement is produced at the contracted level. The team auction must induce the nonpoint sources to truthfully reveal their relative abatement types, providing private abatement cost information. This abatement cost information is needed to select the lowest-cost abatement team, and to set optimal contract prices and quantities for each team member to ensure voluntary participation. Two traditional auction designs will be compared. Efficiency requires an auction that induces truth revelation and allows for incorporation of the optimal abatement strategy under the team contract.

3.1 Introduction

Nonpoint sources are currently the leading cause of water pollution in most areas of the United States (Davies and Mazurek, 1997). Yet, they have avoided intense regulatory scrutiny until fairly recently, due perhaps to the long-standing claim that regulation is impractical because it is inherently difficult to identify individual contributions to nonpoint source pollution loads. The result is that nonpoint sources, such as agriculture, have traditionally been addressed through voluntary subsidy programs that compensate for the adoption of abatement technologies. This tradition has continued in the development of water quality trading (WQT) markets, where nonpoint source participation is voluntary. As a result, WQT markets differ from traditional pollution permit markets, in that point sources are always buyers, and nonpoint sources are always sellers. In this sense, WQT markets are similar to a public procurement or contracting program, with the point source seeking to contract with multiple nonpoint sources to jointly produce a desired level of pollution abatement.

Collective performance-based WQT contracts are complicated by the existence of both adverse selection and moral hazard. However, this is not a problem unique to WQT trading markets. For example, any government that wishes to contract private firms for major public works projects (i.e., construction of highways or dams) must confront similar informational asymmetries. In practice, governments have designed various contracting mechanisms that address both adverse selection (the government does not

know the expected cost of any firm) and moral hazard (the government cannot observe the selected firm's effort to keep its realized production costs low).

What sets the WQT market problem apart is the need to contract jointly with multiple nonpoint sources, coupled with the inability to observe individual productivity. The term "moral hazard in teams" was coined by Holmstrom (1982) to describe this problem. Moral hazard in teams is more pervasive than moral hazard in the single-agent case, as it can occur even when there is no uncertainty in output. Since shirking in effort is only detected through the common final product, the effect of individual shirking is spread across all agents in the group, and cannot be attributed to the responsible party or parties. In this sense, moral hazard in teams is a type of "free-rider" problem prevalent in the provision of public goods.

Even though the actions of the agents are not observable, and so cannot be used as the basis of the contract, the result of the sum of individual actions is verifiable as collective abatement at the end of the period. Therefore, to overcome the moral hazard in teams problem, the collective abatement outcome must be included in the contract that stipulates payment to the agents. A successful contract must pay more when the observable collective performance is a good signal that the individual abatement choices were the required ones. The contract offered by the principal must make each agent feel responsible for the whole of the final product, in order to provide the appropriate incentive for overcoming the free-rider problem.

This paper focuses on the contract design issues associated with the asymmetric information problems inherent in nonpoint source pollution abatement. When a point

source offers a contract for nonpoint source pollution abatement, several informational problems pose obstacles to success. Nonpoint sources have different abilities (i.e., abatement types) in terms of the cost of reducing emissions. The point source cannot observe the expected abatement costs of particular nonpoint sources, and therefore, does not know which firms are the most efficient trading partners. In addition, each nonpoint source is better informed regarding its own unique abatement production process, and therefore holds private information regarding its actual contribution to observed levels of aggregate loadings.

3.2 Existing Literature on Collective Monitoring and Enforcement

Economists have addressed the issue of moral hazard in teams through a wide array of collective monitoring and enforcement mechanisms. Issues of joint production in the labor literature (Holmstrom, 1982; Rasmussen, 1987; McAfee and McMillan, 1991) have been extended to address the joint production problems in nonpoint source pollution control (Meran and Schwalbe, 1987; Segerson, 1988; Xepapadeas, 1991; Cabe and Herriges, 1992; Bystrom and Bromley, 1998, Pushkarskya, 2003). The primary goal of all of these mechanisms is to provide appropriate production incentives to the individual agents, by making each of them liable for the whole team output.

Segerson (1988) applies Holmstrom's (1982) analysis of moral hazard in teams to the problem of nonpoint source pollution, by proposing an ambient tax/subsidy mechanism based on collective monitoring and enforcement. This mechanism pays each

individual a firm specific subsidy, or charges each individual a firm-specific tax, based on the difference between observed levels of aggregate pollution and the collective standard. The team contract that I propose is a member of the same class of budget-breaking collective penalty mechanisms, which is designed for use in a voluntary trading setting. A second major contribution of the propose contract is the use of a team-entry auction to overcome some of the more onerous information requirements of the principal, present in collective performance-based mechanisms.

While collective performance-based instruments are appealing in terms of their theoretical efficiency properties, their adoption as practical policy tools has not occurred. A criticism of collective-performance mechanisms is that they require the principal to possess too much information for efficient implementation in practice. In particular, these mechanisms require the principal, traditionally thought to be a regulatory agency, to have perfect information regarding nonpoint source abatement production and cost functions. It is also expected that nonpoint sources know their own abatement production and cost functions, as well as their impact on aggregate loadings.

Assuming that the required contracting information is privately known by the nonpoint sources, practical implementation of collective performance-based mechanisms requires dealing with an adverse selection problem. Pushkarskya (2003) addresses the role of adverse selection in the design of a nonpoint source pollution abatement subsidy program. The principal (i.e., the regulatory agency) faces both moral hazard in teams and adverse selection. The regulator wishes to target subsidy payments to the nonpoint sources that can produce the most abatement at the least-cost with collective performance

as the only verifiable contract element. A key assumption in this research is that each farmer has perfect information regarding not only their own abatement cost function, but the abatement cost functions of all other farmers. Similar to the Alchian and Demestz (1972) analysis of economic organization, the shared cost information of the nonpoint sources, combined with the potential for economic gains through cooperation, creates an incentive for the nonpoint sources to organize into a trading association. The formation of the trading association circumvents the adverse selection problem, as the point source can contract directly with the association based on observable group abatement. This collective payment can then be divided by the association between individual members using the observable cost information to internally set optimal sharing rules. The contract that I propose in this chapter relaxes the information assumption in Pushkarskya (2003). The contract implements a collective performance-based contract where nonpoint sources only know their own abatement cost function.

3.3 A Model of a Two-stage Team Contract for Water Quality Trading

In this section, a two-stage mechanism that pairs a traditional team contract for the control of moral hazard in teams with an auction to determine team membership prior to contracting is proposed. The contract is designed specifically for application in relatively small watersheds or sub-catchments, where teams can be comprised of a small number of contiguously located neighbors. The task at hand is to design a collective performance-based contract that has the potential for actual implementation. The

contract must transmit to group members clear and readily comprehensible incentives that are consistent with the group's goals of pollution abatement. This means that both incentive compatibility and simplicity, which can be in conflict, are both valued.

In the first stage, multiple nonpoint sources are offered the opportunity to participate in a team contract to produce an aggregate level of abatement, using a sealed bid auction. Bids consist of the quantity of annual pollution reduction the individual pledges to produce, and the corresponding per unit price. The point source accepts the bids of the individuals that, as a group, offer to produce the desired level of nutrient reductions at the lowest cost. The point source will select the group of bidders that, as a team, can reduce the maximum amount of pollution within the point source's fixed budget constraint. The point source's budget constraint is determined by its own abatement cost function. Simply put, the point source will not spend more than it would cost to directly abate the same amount of pollution.

In the second stage, the nonpoint sources will produce abatement. At the end of the second stage, the collective nonpoint source abatement is realized, and participating nonpoint sources are compensated based on the conditional payment schedule. The selected team is then paid according to an "all or nothing" contract based on observed levels of aggregate nonpoint source pollution. If the observed level of collective nonpoint source pollution reductions is greater than or equal to the contracted group level, each team member is paid for their bid quantity at some agreed upon contract

price.[13] However, if the observed level of collective nonpoint source pollution reduction is less than the aggregate quantity, each team member is paid nothing (and suffers a loss equal to the costs of the abatement they produced).

The use of an auction mechanism to select the contracting team members creates incentives for nonpoint sources to truthfully reveal private abatement cost information. This serves two purposes. First, the point source can avoid the adverse selection problem by offering a menu of contracts based on the revealed information. Second, the bid information allows the point source to set optimal sharing rules that ensure that the participation constraints of team members are met. This overcomes the traditional criticism regarding the high informational requirements for efficient team contracts. The role of the team contract is to deliver the appropriate incentives to reduce potential shirking (i.e., abating less than the bid quantity).

This contract allows the point source to purchase a given level of nonpoint source pollution reduction, while leaving nonpoint sources their choice of abatement technology to reduce their pollution discharges. Uncertainty as to the effectiveness of on-site abatement technology is borne by the nonpoint sources, which have better information regarding abatement performance, and are best able to handle it. The competition for team membership effectively limits the range of rent seeking possibilities in nonpoint source bid prices, while endogenously providing the point source with the information needed to set the optimal sharing rules within the contract.

[13] The price is dependent upon the auction design chosen. Under a discriminatory price auction the contract price will be the bid price, and under a uniform price auction the contract price will be equal to the lowest of the non-team members' bid prices. The efficiency and allocative implications of the choice of auction design are addressed in later sections of the chapter.

Consider a watershed that consists of a single, risk neutral point source and multiple, *n*, risk neutral nonpoint sources. A subset of the nonpoint sources, $m \le n$, (i.e., the "team") are selected to provide individually unverifiable levels of abatement, $a_i \in [0, a_i^{\max}]$. Each nonpoint source has a finite capacity for abatement, $a_i^{\max}$. The cost of abatement is given by a strictly increasing convex cost function, $C_i(a_i)$. The team's aggregate abatement, $A(a,e)$, depends stochastically on the individual abatement actions of the nonpoint sources and random weather effects. Expected team abatement is denoted as: $E[A(a,e)] = \omega \sum_{j=1}^{m} a_i(1+e) + (1-\omega)\sum_{j=1}^{m} a_i(1-e)$, where ω and $1-\omega$ are the probabilities of good and bad weather respectively $(0 < \omega < 1)$, and $1+e$ and $1-e$ represent the impact of good and bad weather respectively $(0 < e < 1)$. Any increase in individual abatement increases the expected level of team abatement, i.e., $E\left[\frac{\partial A(a,e)}{\partial a_i}\right] > 0$. The principal offers an "all-or-nothing" team contract that makes individual payments contingent on the monitoring of team performance. The team target Λ is the sum of the individually contracted quantity of abatement for all team members, $\sum_{j=1}^{m} \lambda_j = \Lambda$. If the observed level of aggregate nonpoint source pollution abatement is greater than or equal to the team target, individual team members receive a positive payment. However, if aggregate nonpoint source pollution abatement is less than the team target, payments are withheld from each team member. Appendix A contains a table defining the notations used throughout the chapter.

3.4 A Budget-breaking Team Contract to Avoid Moral Hazard in Teams

As mentioned previously, the proposed WQT contract must confront the combined effects of adverse selection and moral hazard on WQT contract design. However, I will begin with an analysis of moral hazard in teams, in isolation, before including adverse selection. This allows for the effects of the team contract and the team entry auction to be examined separately. Within this section, it is assumed that the point source knows the efficiency type of each nonpoint source (i.e., the point source knows the abatement cost functions of all nonpoint sources). This allows the point source to select the optimal combination of low-cost trading partners and the optimal sharing rules, without the auction.

The point source offers individualized contracts of the following type:

$r_i = \begin{cases} p_i \lambda_i & \text{if } \mathrm{A(a,e)} \geq \Lambda \\ 0 & \text{if } \mathrm{A(a,e)} < \Lambda \end{cases}$. The symbols p_i and λ_i represents the price paid per unit of abatement to nonpoint source *i*, and the quantity of abatement contracted from nonpoint source *i*, respectively. The profit for each of the *m* nonpoint sources under contract can also be represented as: $\pi_i = \begin{cases} p_i \lambda_i - C_i(a_i) & \text{if } \mathrm{A(a,e)} \geq \Lambda \\ -C_i(a_i) & \text{if } \mathrm{A(a,e)} < \Lambda \end{cases}$. Therefore, the point source can set p_i and λ_i, such that when the target is achieved each team member earns profit, and when the target is not met each team member suffers a loss. Because nonpoint source market participation is voluntary, the point source must be concerned with meeting each individual's participation constraint. The point source in this case

must set the contract price and quantity in order to ensure that each nonpoint source will be made no worse off by participating optimally within the contract.

The participation constraint requires that the payment to each nonpoint source must be greater than or equal to their actual costs of abatement, i.e., $p_i \lambda_i \geq C_i(a_i^*)$, which can be rewritten as $p_i \geq \frac{C_i(a_i^*)}{\lambda_i}$. The right hand side of this inequality is an "adjusted" average abatement cost. Since the point source does not observe the individual abatement decision of the nonpoint sources, actual abatement, a_i, can be greater than, less than, or equal to the contract quantity, λ_i. Thus, the break-even condition of price greater than average cost must be adjusted to take the potential discrepancy between actual and contracted quantities of abatement. The point source will offer a price that guarantees the nonpoint source will, at a minimum, breakeven when producing the desired level of abatement. Under voluntary participation, the breakeven condition is identical to the participation condition. The concept of adjusted average cost is used throughout the remainder of the chapter for this reason.

In order to explicitly include the effects of weather on the profit maximizing decision of the nonpoint source, I write the expected profit function:

$$\underset{a_i}{Max} \quad \alpha + \beta$$

where,

$$\alpha \equiv \omega \left[\left(p_i \lambda_i\right) I\left((1+e)\left(\sum_{j \neq i} a_j + a_i \right) \geq \Lambda \right) - C_i\left(a_i\right) \right]$$

$$\beta \equiv (1-\omega) \left[\left(p_i \lambda_i\right) I\left((1-e)\left(\sum_{j \neq i} a_j + a_i \right) \geq \Lambda \right) - C_i\left(a_i\right) \right].$$

This represents the expected profit of nonpoint source *i*, given the choice of abatement level, a_i. The conditional payment structure of the team contract is captured through the use of the indicator function, $I(\cdot)$. When the condition inside the indicator function holds, total observed abatement meets or exceeds the group target, and the value of the indicator function is equal to one. Otherwise, the value of the indicator function is equal to zero. The first portion of the objective function, α, represents the expected profit in good weather conditions. Good weather occurs with probability ω, and increases the level of collective abatement observed downstream by $(1+e)$. The second portion of the objective function, β, represents the expected profit in bad weather conditions. Bad weather occurs with probability $(1-\omega)$, and decreases the level of collective abatement observed downstream by $(1-e)$.

The Nash equilibrium abatement strategy maximizes the expected profit of each nonpoint source. The discrete weather distribution assumption requires analysis of the abatement decision given three weather/payment contingent scenarios (Table 1).[14]

	Good Weather	**Bad Weather**
1	*Collective target met:* $I(\cdot)=1$	*Collective target met:* $I(\cdot)=1$
2	*Collective target met:* $I(\cdot)=1$	*Collective target not met:* $I(\cdot)=0$
3	*Collective target not met:* $I(\cdot)=0$	*Collective target not met:* $I(\cdot)=0$

Table 3.1: Weather/Payment Contingent Scenarios

[14] We do not consider the combination where there is a positive payout in bad weather and a zero payment in good weather as this is not feasible using a single abatement strategy.

Within the first scenario the nonpoint sources choose the optimal abatement strategy that ensures a payment in both weather states. The second scenario has the nonpoint sources choose the optimal abatement strategy to ensure payment is met only in the event of good weather. The third scenario has the nonpoint sources choose to miss the target and forgo payment in both states of weather. Once the optimal strategies under each scenario are determined the Nash abatement strategy can be selected from among them.

Assume that the point source sets the optimal sharing rules (p_i , λ_i) such that the participation constraint of each team member will be met, $p_i \geq \frac{C_i(a_i^*)}{\lambda_i}$. With these assumptions in place, it will be shown that the abatement production strategy that maximizes expected profit, contingent upon meeting the target in both good and bad weather, is $a_i = \frac{\lambda_i}{(1-e)}$. To show this, let $a_j = \frac{\lambda_j}{(1-e)}$ be the abatement production strategy for all nonpoint sources $j \neq i$, who were selected into the team. Nonpoint source i's profit maximizing choice of abatement is determined by maximizing expected utility:

$$\underset{a_i}{Max} \qquad (\alpha) + (\beta)$$

$$where \quad (\alpha) \equiv \omega\left[\left(p_i\,\lambda_i\right) I\left((1+e)\left(\left(\frac{1}{(1-e)}\sum_{j\neq i}\lambda_j\right) + a_i\right) \geq \Lambda\right) - C_i\left(a_i\right)\right] \quad . \quad \text{Eq (3.1)}$$

$$and \quad (\beta) \equiv (1-\omega)\left[\left(p_i\,\lambda_i\right) I\left((1-e)\left(\left(\frac{1}{(1-e)}\sum_{j\neq i}\lambda_j\right) + a_i\right) \geq \Lambda\right) - C_i\left(a_i\right)\right]$$

It has been assumed that the nonpoint sources are choosing the expected profit maximizing abatement strategy that will also guarantee payment in both good and bad

weather. This is equivalent to choosing the expected profit maximizing level of abatement that guarantees $I(\cdot)=1$ in both *(α)* and *(β* of Eq. (3.1).

Solving both indicator functions provides the minimal abatement levels at which the collective target will be met in each weather period, given the assumption regarding the abatement of the other nonpoint sources. Solving the contents of indicator function from part (α) of Eq. (3.1):

$$
\begin{aligned}
&(1+e)\left(\left(\frac{1}{(1-e)}\sum_{j\neq i}\lambda_j\right)+a_i\right)\geq\Lambda\\
&=\left(\frac{(1+e)}{(1-e)}\sum_{j\neq i}\lambda_j\right)+(1+e)a_i\geq\left(\sum_{j\neq i}\lambda_j\right)+\lambda_i\\
&=(1+e)a_i\geq\left(\sum_{j\neq i}\lambda_j\right)+\lambda_i-\left(\frac{(1+e)}{(1-e)}\sum_{j\neq i}\lambda_j\right)\\
&=a_i\geq\sum_{j\neq i}\lambda_j\left(\frac{1}{(1+e)}-\frac{1}{(1-e)}\right)+\frac{\lambda_i}{(1+e)}
\end{aligned}
$$

Solving the indicator function from part (β) of Eq. (3.1):

$$
\begin{aligned}
&(1-e)\left(\left(\frac{1}{(1-e)}\sum_{j\neq i}\lambda_j\right)+a_i\right)\geq\Lambda\\
&=\left(\frac{(1-e)}{(1-e)}\sum_{j\neq i}\lambda_j\right)+(1-e)a_i\geq\left(\sum_{j\neq i}\lambda_j\right)+\lambda_i\\
&=(1-e)a_i\geq+\lambda_i\\
&=a_i\geq\frac{\lambda_i}{(1-e)}
\end{aligned}
$$

Since we wish to achieve the collective target in both periods, we need to choose the abatement strategy that produces the larger amount of abatement. To do this, it must

be determined whether $\frac{\lambda_i}{(1-e)}$ produces more abatement than

$$\sum_{j \neq i} \lambda_j \left(\frac{1}{(1+e)} - \frac{1}{(1-e)} \right) + \frac{\lambda_i}{(1+e)} :$$

$$\frac{\lambda_i}{(1-e)} \geq \sum_{j \neq i} \lambda_j \left(\frac{1}{(1+e)} - \frac{1}{(1-e)} \right) + \frac{\lambda_i}{(1+e)}$$

$$= \frac{\lambda_i}{(1-e)} \geq \frac{\Lambda}{(1+e)} - \frac{\sum_{j \neq i} \lambda_j}{(1-e)}$$

$$= \frac{\Lambda}{(1-e)} \geq \frac{\Lambda}{(1+e)}$$

Therefore, $a_i \geq \frac{\lambda_i}{(1-e)}$ is the optimal, contingent on receiving payment in both weather conditions.

Finally, it must be shown that the abatement strategy actually holds at equality, $a_i = \frac{\lambda_i}{(1-e)}$, by maximizing expected profits, subject to the chosen abatement strategy.

$$\underset{a_i}{Max} \qquad \omega\left(p_i \lambda_i (1) - C_i(a_i)\right) + (1-\omega)\left(p_i \lambda_i (1) - C_i(a_i)\right)$$

$$s.t. \qquad a_i \geq \frac{\lambda_i}{(1-e)}$$

Construct the Hamiltonian:

$$\underset{a_i}{Max} \qquad H = p_i \lambda_i - C_i(a_i) + k_1 (a_i - \frac{\lambda_i}{(1-e)})$$

$$\frac{\partial H}{\partial a_i} \equiv \frac{\partial C_i(a_i)}{\partial a_i} = k_1$$

$$\begin{cases} k_1 > 0 \\ a_i = \dfrac{\lambda_i}{(1-e)} \end{cases} \qquad \begin{cases} k_1 = 0 \\ a_i > \dfrac{\lambda_i}{(1-e)} \end{cases}$$

The slack condition, $k_1 > 0$, holds with inequality, since $\frac{\partial C_i(a_i)}{\partial a_i} > 0$. Therefore, the abatement strategy holds with equality, $a_i^* = \frac{\lambda_i}{(1-e)}$. The rationale being that at equality the abatement production strategy ensures payment in both good and bad weather. Any additional abatement will increase costs but will not change the expected revenue.

In order to determine the remaining feasible abatement production strategies given this group contract, the same process needs to be repeated to determine the optimal abatement strategies for the two remaining weather/payment contingent scenarios. When the nonpoint sources only wish to achieve the target in the event of good weather, the optimal abatement strategy is $a_i = \frac{\lambda_i}{(1+e)}$. When the nonpoint source does not wish to achieve the target in either period, the optimal abatement strategy is $a_i = 0$. For the sake of brevity, the calculations of these optimal abatement strategies have been relegated to Appendix B and C, respectively. The optimal weather/payment contingent abatement strategies can be labeled as:

(i) *the surplus strategy,* $a_i(1-e) = \lambda_i$;

(ii) *the shortfall strategy,* $a_i(1+e) = \lambda_i$; and,

(iii) *the non-participating strategy,* $a_i = 0$.

The expected profit maximizing strategy among these choices is determined by the distribution of weather, the size of the weather shock, and the contract price and quantities.

The nonpoint source that follows the surplus strategy will produce more than its bid quantity of abatement, so that the target is achieved regardless of weather. The amount of overproduction equals the expected abatement shortfall that occurs under bad weather. Expected profit under the surplus strategy is:

$$E[\pi_i]_{Surplus} = \omega\left(p_i a_i (1-e) - C_i(a_i)\right) + (1-\omega)\left(p_i a_i (1-e) - C_i(a_i)\right) = p_i a_i (1-e) - C_i(a_i)$$

The nonpoint source that follows the shortfall strategy will produce less than its bid quantity of abatement, so that the aggregate abatement target is reached only in the event of good weather. The amount of underproduction equals the expected level of the abatement windfall under good weather. Thus, the amount of the shortfall is constrained by the weather effect. To avoid a loss the nonpoint source must produce enough to guarantee a payment in good weather. Expected profit under the shortfall strategy is:

$$E[\pi_i]_{Shortfall} = \omega\left(p_i a_i (1+e) - C_i(a_i)\right) + (1-\omega)\left(-C_i(a_i)\right) = \omega\left(p_i a_i (1+e)\right) - C_i(a_i).$$

The non-participation strategy is the optimal abatement strategy when the target will not be achieved regardless of the weather. In this situation, the expected revenue is equal to zero. The optimal response is to not produce any abatement, and, thus, not incur any costs. The expected profit under the non-participating strategy is:

$$E[\pi_i]_{Non-participation} = \omega\left(-C_i(0)\right) + (1-\omega)\left(-C_i(0)\right) = 0.$$

It is clear that the non-participating strategy cannot be the optimal strategy, when the participation constraint is met, since both the surplus and shortfall strategy return positive levels of expected profit. Therefore, the Nash equilibrium abatement strategy will be either surplus or shortfall determined by the difference in expected profit:

$$E_{surplus}[\pi_i] > E_{shortfall}[\pi_i]$$
$$p_i a_i (1-e) - C_i(a_i) > \omega(p_i a_i (1+e)) - C_i(a_i)$$
$$(1-e) > \omega(1+e)$$

When this condition holds, the surplus strategy returns higher expected profits than the shortfall strategy, and vice versa. The Nash equilibrium abatement strategy is determined by the probability of good weather, ω, and the magnitude of the weather shock, *e*. The surplus strategy, $a_i(1-e) = \lambda_i$, is the Nash equilibrium if $(1-e) > \omega(1+e)$, and the shortfall strategy, $a_i(1+e) = \lambda_i$, is the Nash equilibrium if $(1-e) < \omega(1+e)$.

These probabilities are common knowledge to the point source and all the nonpoint sources, and thus the Nash abatement strategy will be known as well. This allows the point source to assign each team member's contract price and quantity in orderto maximize its own expected abatement cost savings. Ideally, the point source would like to set the contract prices and quantities as in Figure 3.1. The contract price is set equal to the point source's own marginal cost of abatement at the aggregate level of nonpoint source abatement being purchased. This point is denoted as β in Figure 3.1. In addition, the point source would choose to set the contract quantities of each team member equal to their economically efficient production levels, $\lambda_i = a_i^*$.

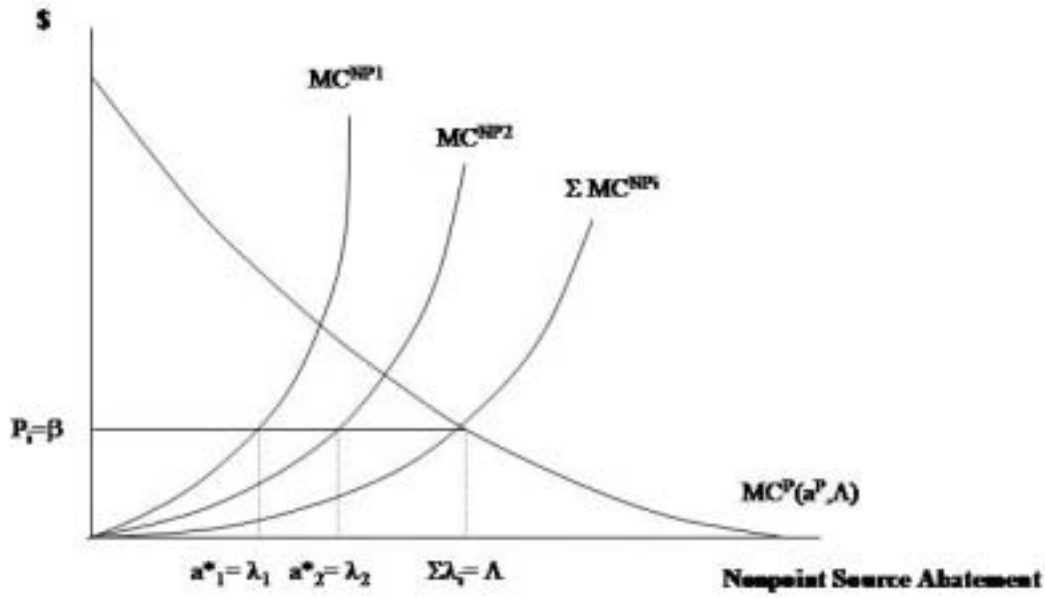

Figure 3.1: First-Best Allocation of Contract Price and Quantity

However, from the previous analysis it is clear that the team contract will not produce this result. If the Nash abatement strategy is *shortfall*, the nonpoint source will under-produce abatement, and if the Nash abatement strategy is *surplus* the nonpoint source will overproduce abatement. The point source must deviate from the first-best contract price in order to guarantee the contracted level of individual abatement is produced by each nonpoint source. This is the common second-best result attributable to asymmetric information (Sandmo).

If the structure of probable weather shocks are such that $(1-e) > \omega(1+e)$, then the point source and the nonpoint sources all know that the surplus abatement strategy is the optimal strategy to follow. The point source will set the contract quantity at the first-best efficiency level $\lambda_i = a_i^* |_\beta$ and $p_i = \dfrac{\beta}{(1-e)}$ to ensure that the nonpoint source produces abatement $a_i^* = \lambda_i$. To ensure that the desired level of abatement is provided by each nonpoint source, the point source will offer a contract price that is adjusted to account for the abatement strategy that the nonpoint sources will optimally follow. The nonpoint source will choose the surplus abatement strategy $a_i(1-e) = \lambda_i$, solving the expected profit maximization problem:

$$\begin{aligned} &\underset{a_i}{Max} \quad p_i\lambda_i - C_i(a_i) \\ &s.t. \quad a_i(1-e) = \lambda_i \end{aligned} \quad \forall i$$

The first order conditions are:

$$p_i(1-e) = \frac{\partial C_i(a_i)}{\partial a_i} \quad \forall i$$

Setting $p_i = \frac{\beta}{(1-e)}$ results in $\beta = \frac{\partial C_i(a_i^*)}{\partial a_i^*}$ which has the nonpoint source producing the desired level of abatement $a_i^* = \lambda_i$.

Alternatively, if the structure of probable weather shocks are such that $(1-e) < \omega(1+e)$, then the point source and the nonpoint sources all know that the shortfall abatement strategy is the optimal strategy to follow. The point source will again set the contract quantity at the first-best efficiency level $\lambda_i = a_i^* |_\beta$, but will now set the contract price as $p_i = \frac{\beta}{\omega(1+e)}$ to ensure that the nonpoint source produces abatement $a_i^* = \lambda_i$.

The nonpoint source will choose the shortfall abatement strategy $a_i(1+e) = \lambda_i$, solving the expected profit maximization problem:

$$\begin{aligned} &\underset{a_i}{Max} \quad \omega p_i \lambda_i - C_i(a_i) \\ &s.t. \quad a_i(1+e) = \lambda_i \end{aligned} \quad \forall i$$

The first order conditions are:

$$\omega p_i (1+e) = \frac{\partial C_i(a_i)}{\partial a_i} \quad \forall i$$

Setting $p_i = \frac{\beta}{\omega(1+e)}$ results in $\beta = \frac{\partial C_i(a_i^*)}{\partial a_i^*}$ which has the nonpoint source producing the desired level of abatement $a_i^* = \lambda_i$.

In either case, the point source must set the contract price higher than the first best level in order to ensure that the desired level of individual nonpoint source abatement is

provided from each team member. Figures 3.2 and 3.3 illustrate this for the surplus and shortfall abatement strategies respectively. In both cases, some level of informational rent is extracted by the nonpoint source due to the presence of asymmetric information, and the Nash abatement strategy is the one that provides the largest information rent to the nonpoint sources.

These results, not surprisingly, are similar to those of the collective mechanisms proposed by Holmstrom (1982) and Segerson (1988). This is because the team contract proposed in this dissertation belongs to the same class of budget-breaking collective mechanisms. Some manipulation of terms can illustrate the commonality of this contract with that of Holmstrom and Segerson. Theorem 3 in Holmstrom (1988) proposes the contract:

$$s_i(x) = \begin{cases} s_i x & x \geq \bar{x} \\ s_i x - k_i & x < \bar{x} \end{cases}$$

where, s_i is the share of the collective output value (x) attributed to team member i, with $\sum_i s_i = 1$, and $k_i > 0$ is a fine for failing to achieve the contract level $(\bar{x})$. The WQT contract presented in this chapter can be rewritten to correspond with this Holmstrom contract as follows:

$$s_i(x) = \begin{cases} s_i \bar{x} & x \geq \bar{x} \\ s_i \bar{x} - k_i & x < \bar{x} \end{cases}$$

where, $s_i = \dfrac{\lambda_i}{\Lambda}$ and $\bar{x} = p\Lambda$. The differences between the two contracts are in terms of $\bar{x}$ and k_i. In this contract the output value is fixed at the contracted level $\bar{x}$. In addition,

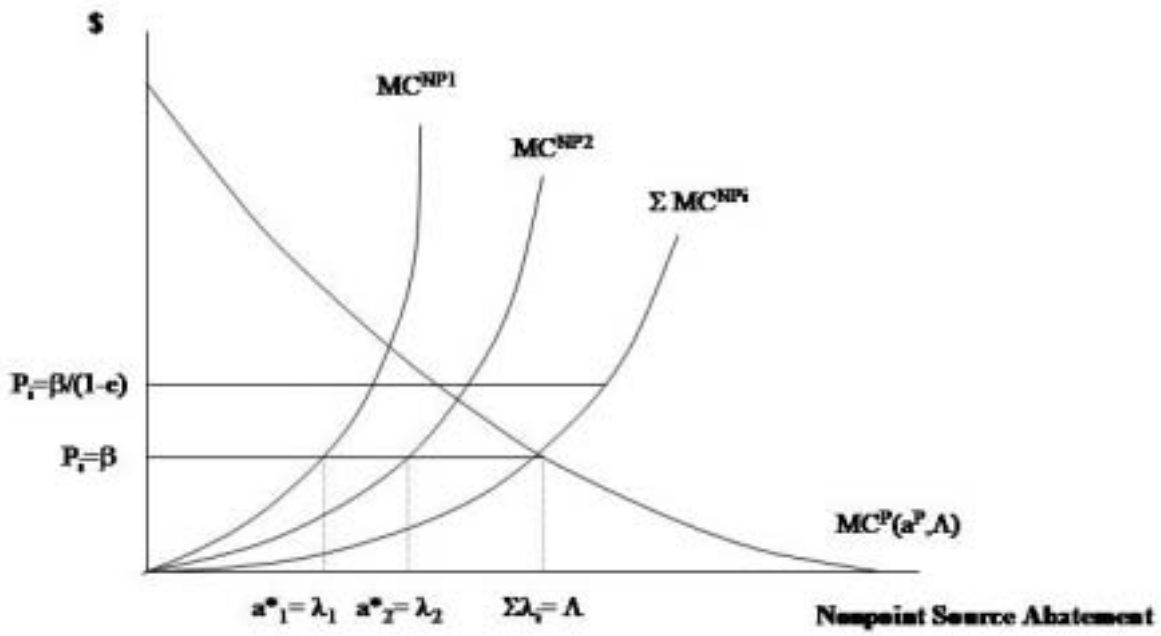

Figure 3.2: Second-Best Allocation of Contract Quantity and Price (Surplus Abatement Strategy and No Adverse Selection)

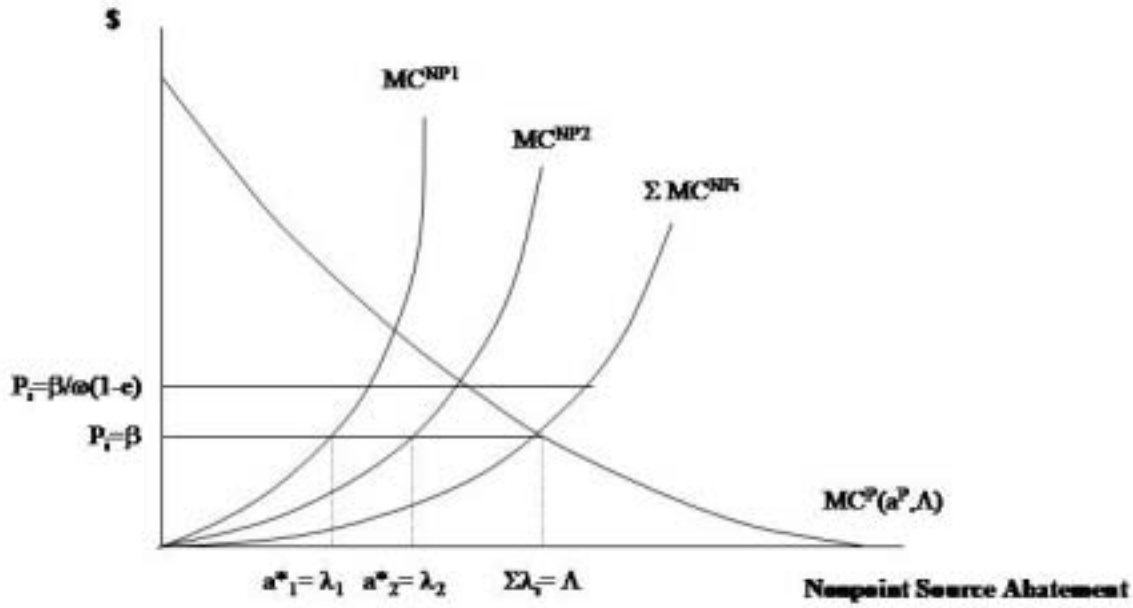

Figure 3.3: Second-Best Allocation of Contract Quantity and Price (Shortfall Abatement Strategy and No Adverse Selection)

the fine for under-compliance, k_i, is always set equal to the individual's share of the fixed output value ($k_i = s_i \bar{x}$). The principal seeking full compliance in the face of production uncertainty, will not seek to adjust the fine, as in Holmstrom and Segerson, but instead will adjust the optimal value for output, p_i.[15]

A common problem often discussed in association with both the Holmstrom and Segerson instruments are the effects of endowment constraints. Endowment constraints restrict the credibility of many collective performance instruments. Under certain conditions, the size of fines required for full compliance may be so large as to eclipse the wealth of the individual agents. In such settings, the mechanism loses its practical enforcement credibility (Karp, 2002). The same is true in the WQT setting, where the per unit price required to secure individual contract quantities may be prohibitively high, especially when weather impacts are large, or when the difference in marginal abatement costs between point and nonpoint sources is small. The budget constraint of the point source provides a default to nonpoint source trading. The point source can always opt to produce its own abatement at cost. When the cost of contracting with nonpoint sources is too high, the budget constraint will be exceeded, and the proposed contract will not permit trading. Unlike the Segerson and Holmstrom mechanisms, endowment constraints do not weaken the credibility of the contract incentives, rather, they preclude trade from occurring at all.

[15] Changing the output value does affect the size of the fine, as it is the fine is equivalent to the loss of the individual share of total output value under non-compliance. However, this is a direct result of the change in contract price.

3.5 An Auction Mechanism to Determine Team Entry and Sharing Rules

When the assumption regarding the knowledge of individual nonpoint source abatement cost functions is relaxed, the point source no longer possesses the information needed to select efficient team members and set optimal sharing rules. In other words, the point source is no longer able to identify the lowest-cost subset of nonpoint source polluters. In addition, the point source does not have the information needed to set the optimal contract price such that the participation constraint and the production of the contract quantity of abatement is ensured. Introducing adverse selection requires the use of a mechanism that can induce the nonpoint sources to voluntarily reveal this private information. Auctions are commonly used for this purpose, and among the wide array of auction designs, two types have generally received the most attention: uniform price and discriminating price, sealed-bid auctions (Harris and Raviv, 1981). In a uniform price auction, a single market price equal to the lowest rejected bid is paid to all accepted bidders. In a discriminating auction, all winning bidders are paid their bid price, rather than a common price.

The WQT setting presents some unique challenges regarding the appropriate choice of the auction mechanism. Research on multi-unit auctions has shown that the efficiency conditions proven in the single-unit case are not guaranteed to hold when individual bidders offer multiple-units. In particular, bid-shading in both uniform price and discriminating price auctions leads to inefficient allocation when bidders are allowed to bid for multiple units at differing prices, or when any bidder(s) can exert market power

(Ausubel and Cramton, 2002). In the uniform case, the ability to bid for multiple quantities at differing prices can lead to a bidder's own bid being "pivotal" in setting the price, and thus, truth revealing bids are no longer a dominant strategy. By offering multiple units at varying prices, the bidder can potentially be selected into the team and also set the market price, which creates an incentive to shade bid prices on some units. Market power allows the dominant bidder to influence the auction price, in the discriminating price auction.

To avoid the potential inefficiencies attributed to multi-unit auctions, I assume that each nonpoint source has the same maximum capacity for abatement: $a_i^{\max} = a_j^{\max} = a^{\max}$, and thus will have a maximum, strategy contingent, bid quantity, $\lambda_i^{\max} = \lambda_j^{\max} = \lambda^{\max}$. This assumption removes the potential for market domination by any single bidder. In addition, I restrict bidders to a single bid price over all bid quantities. This prevents any bidder's bid from being pivotal in determining its own price in the uniform auction. Under these assumptions, the inefficiencies of multi-unit auctions are avoided (Ausubel and Cramton, 2002).

In the team entry auction, a single risk-neutral point source desires to purchase Λ amount of nonpoint source abatement, subject to a budget constraint. The budget constraint is determined by the point sources own abatement cost function (i.e., the point source will not pay more than its own cost for abatement). Multiple risk neutral nonpoint sources, *n*, bid to join the abatement team. The $m \leq n$ nonpoint sources that collectively produce Λ level of abatement at the lowest-cost are chosen for team entry. Bids consist

of a per unit price, $p_i(\lambda_i)$, for a quantity of abatement, λ_i. Ties are broken through random selection.

Following Wilson (1979) and Ausubel and Cramton (2002), I represent the multi-unit auction in terms of shares, by normalizing the collective abatement target Λ=1, with individual quantity bids $\lambda_i \in (0, \lambda^{\max})$, and $\lambda^{\max} \in (0,1)$. This simplifies the construction of order statistics needed for an analytical solution for optimal bidding strategies.

Order statistics are useful tools in the analysis of auctions. The point source and all rival nonpoint sources assume that the reservation prices of all bidders $(\theta_1,...,\theta_n)$ are identical independently drawn random variables from a cumulative density function $G(\cdot)$ with probability density function, $g(\cdot)$. The distribution of reservation prices is known by all, but the individual realization θ_i is only known to bidder *i*.

By arranging the *n* i.i.d. random reservation prices in ascending order of magnitude $(\theta_{(1)} \le \theta_{(2)} \le ... \le \theta_{(n)})$, we can denote the mth order statistic of all bidders other than *i* as $\theta_{(m)}^{-i}$. Now we can denote $F_{(m)}^{-i}$ as the cumulative density function and $f_{(m)}^{-i}$ the probability density function of the mth order statistic. The probability of being accepted into the team can be written as a function of the distribution of the mth order statistic, $\Pr[p_i(\lambda_i) \;\; accepted] = \Pr[p_i(\lambda_i) < \theta_{(m)}^{-i}] = 1 - F_{(m)}^{-i}$.

In the following sections, the same method is used to determine the optimal bid price and quantities for both possible Nash equilibrium abatement strategies. To avoid repetition I will only report the details for the surplus strategy within the text. The

derivation of the optimal bid price and quantity under the shortfall abatement strategy is identical to that presented, and can be found in Appendix D (Uniform Price Auction) and Appendix E (Discriminating Price Auction).[16]

3.6 Uniform Price Team Entry Auction

In a uniform price auction the bidder faces uncertainty in regard to both team entry and the contract price. The *m* lowest-priced bidders, that collectively bid to produce the team target ($\sum_{i}^{m} \lambda_i = \Lambda$), are each contracted to produce their bid quantity at a per unit price equal to the lowest rejected bid. The bid price of the lowest excluded bidder (*m+1*) becomes the contract price, referred to as the "stop-out" price, for all team members. The "stop-out price" ($\overline{p}_{m+1}$) is an expected price, over the distribution of the order statistic distribution, $1 - F_{(m)}^{-i}$. As will be shown, it can also be interpreted as the expected average cost of the *m+1* bidder, $\frac{C_{m+1}(a_{m+1})}{\lambda_{m+1}}$.

Each nonpoint source will bid its true reservation price in the uniform price auction. Truth revelation implies that the bid will reflect the actual costs of abatement for the nonpoint source: $p_i(a_i^*)a_i^* = C_i\left(a_i^*\right)$. Therefore, the nonpoint source bid price holds

[16] The only difference between the surplus and shirking abatement strategies is in terms of the determination of optimal bid price and quantity is in the weather shock term. Under the shirking abatement strategy ω(1+e) is used in place of (1-e).

the following relationship to the reservation price: $p_i(a_i^*) = \frac{C_i\left(a_i^*\right)}{a_i^*} = \theta_i$. The uniform price auction ensures that the individual rationality constraints of all accepted team members are met, as the contract price is always greater than the bid price of the team members.

The competitive team entry auction is a revelation mechanism, which induces the nonpoint source to bid a price that equals its average abatement cost: $p_i(a_i^*) = \frac{C_i\left(a_i^*\right)}{a_i^*} = \theta_i$. This is because any deviation from the truth revealing bidding strategy will not improve the expected outcome for the bidder. This is true for either increasing or decreasing the bid price relative to the reservation price.

When a bidder sets the bid price $p_i(\lambda_i)$ lower than the reservation price θ_i, three possible scenarios exist:

$$\begin{aligned} &(a) \quad p_i(\lambda_i) < \theta_i < \overline{p}_{m+1} \\ &(b) \quad \overline{p}_{m+1} < p_i(\lambda_i) < \theta_i \quad or \\ &(c) \quad p_i(\lambda_i) \le \overline{p}_{m+1} < \theta_i . \end{aligned}$$

When the nonpoint source's reservation price, θ_i, is less than the expected competitive auction price, $\overline{p}_{m+1}$, as in scenario *(a)*, decreasing the bid price $p_i(\lambda_i)$ has no effect. The nonpoint source stays in the team and receives the same competitive price. When ϕ_i is greater than $\overline{p}_{m+1}$, reducing the bid price can only make the nonpoint source worse off. Reducing $p_i(\lambda_i)$ to any point greater than $\overline{p}_{m+1}$, as in scenario *(b)*, does not gain the nonpoint source entry into the team, and its auction outcome is unchanged. Setting $p_i(\lambda_i)$ equal to or less than $\overline{p}_{m+1}$, as in scenario *(c)*, worsens the auction outcome of the

nonpoint source. When $p_i(\lambda_i) = \overline{p}_{m+1}$, the nonpoint source has a random chance of being selected into the team. If the nonpoint source does not gain entry into the team his status remains unchanged, and if selected the nonpoint source is guaranteed a loss because the auction price will be less than the average cost of abatement. Reducing $p_i(\lambda_i)$ below $\overline{p}_{m+1}$ exacerbates this loss.

When a bidder sets the bid price higher than the reservation price, three possible scenarios exist:

$$\begin{aligned} &(c) \quad \overline{p}_{m+1} < \theta_i < p_i(\lambda_i) \\ &(d) \quad \theta_i < p_i(\lambda_i) < \overline{p}_{m+1} \quad \textit{or} \\ &(e) \quad \theta_i < \overline{p}_{m+1} \leq p_i(\lambda_i). \end{aligned}$$

When the reservation price is greater than the stop-out price, as in scenario (*c)*, increasing the bid price does not affect the auction outcome as the nonpoint source will continue to remain outside of the team. When the reservation price is less than the stop-out price, the truth revealing bidding strategy ensures the nonpoint source entrance to the team. Increasing the bid price to any point less than the stop-out price, as in scenario *(d)*, has no effect on the auction outcome. The nonpoint source remains in the team and will receive per unit stop-out price. Raising the bid price equal to or greater than stop-out price, as in situation *(e)* puts the nonpoint source in danger of suffering a loss. Bidding the stop-price results in the nonpoint source having a random chance of being selected into the team. If the nonpoint source is selected into the team the auction outcome remains the same, (i.e., he is paid the stop-price). However, if not selected the nonpoint source suffers the loss of expected profit he would have made had he remained in the team. Obviously, the same loss occurs for any bid price set above the stop-price. It is a dominant strategy for the

nonpoint source to bid the reservation price associated with the chosen abatement strategy.

The optimal bid price line is the average abatement cost curve., i.e., $p(a_i^*) = \frac{C_i(a_i^*)}{a_i^*}$. Thus, the optimal bid price is simply the average cost of abatement at the optimal quantity level. In a uniform price auction, the nonpoint source will always bid the feasible quantity of abatement that maximizes expected profits, given the Nash abatement strategy.

The utility maximizing quantity of abatement under the surplus abatement strategy is determined by solving the following expected profit maximization problem:

$$\begin{aligned} &\underset{a_i,\lambda_i}{Max} \quad \overline{p}_{m+1}\lambda_i - C_i(a_i) \\ &s.t. \qquad \lambda_i = a_i(1-e) \\ &\qquad\quad \lambda^{\max} = a^{\max}(1-e) \\ &\qquad\quad \lambda_i \le \lambda^{,\max} \end{aligned}$$

This maximization problem can be simplified as:

$$\begin{aligned} &\underset{a_i}{Max} \quad \overline{p}_{m+1}a_i(1-e) - C_i(a_i) \\ &s.t. \qquad a_i = a^{\max}(1-e) \end{aligned}$$

$$H = \overline{p}_{m+1}a_i(1-e) - C_i(a_i) + k_1(a^{\max}(1-e) - a_i)$$

Show first order conditions with slack constraints:

$$\frac{\partial H}{\partial a_i} \equiv \overline{p}_{m+1} - \frac{\partial C_i(a_i)}{\partial a_i}\frac{1}{(1-e)} - k_1 = 0$$

$$\begin{cases} k_1 > 0 \\ a_i = a^{\max} \end{cases} \qquad or \qquad \begin{cases} k_1 = 0 \\ a_i < a^{\max} \end{cases}$$

The slack conditions can be interpreted in the following manner. If the expected contract price is greater than the marginal cost of abatement, which is adjusted to account for the Nash abatement strategy, at the maximum allowable bid quantity ($a^{\max}$), the nonpoint maximizes expected profit by bidding this amount. Otherwise, the nonpoint source will bid the quantity(a_i^*) where the "stop-out" price equals the adjusted marginal cost of abatement. The optimal bid quantity is the expected profit maximizing level: $\lambda_i^* \min\left[a^{\max}, a_i^*\right]$.

The uniform price auction serves as a truth revelation mechanism, as the optimal bid price is equal to the average cost of abatement at the optimal bid quantity. The optimal bid quantity is less than the first-best profit maximizing level as a result of informational rents from asymmetric information. The nonpoint source accounts for its abatement strategy of over-production in its selection of its bid quantity. However, the contract does ensure that each team member will produce individual abatement levels equal to their bid quantities. The optimal bidding strategy is depicted graphically in Figure 3.4.

3.7 Discriminating Price Team Entry Auction

In the discriminating auction, the bidder faces uncertainty about acceptance, but not about price. The *m* lowest-priced bidders, that collectively bid to produce the team target ($\sum_i^m \lambda_i = \Lambda$), are each contracted to produce their bid quantity at their bid price.

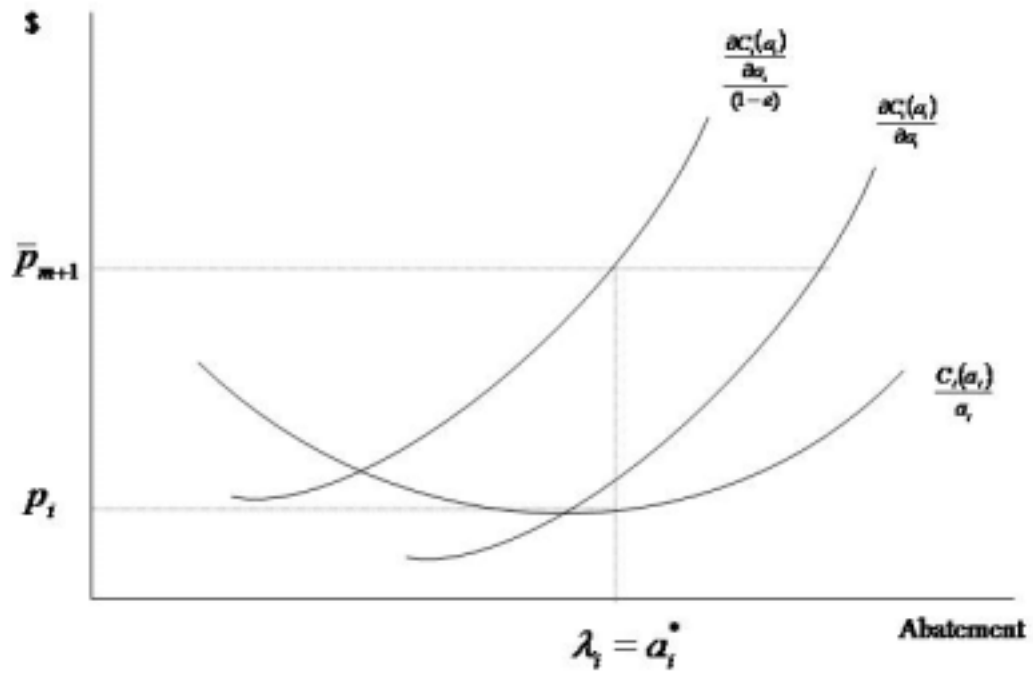

Figure 3.4: Optimal Bid Price and Quantity (Uniform Price Auction and Surplus Abatement Strategy)

Thus, all bidders have an incentive to increase their bid price above the reservation price. In a discriminating price auction the nonpoint source maximizes expected profit, where the expectation is now based on the uncertainty associated with being selected into the team[17], represented by the distribution of the mth order statistic:

$$\underset{p_i, a_i, \lambda_i}{Max} \left(1 - F_{(m)}^{-i}\right)\left(p_i \lambda_i - C_i\left(a_i\right)\right) + F_{(m)}^{-i}\left(0\right)$$

$$s.t. \qquad a_i = \frac{\lambda_i}{(1-e)}$$

$$\lambda_i \leq \lambda^{\max}$$

$$a^{\max}(1-e) = \lambda^{\max}$$

The maximization problem can be simplified as:

$$\underset{p_i, a_i, \lambda_i}{Max} \left(1 - F_{(m)}^{-i}\right)\left(p_i \lambda_i - C_i\left(a_i\right)\right) + F_{(m)}^{-i}\left(0\right)$$

$$s.t. \qquad a_i(1-e) = \lambda_i$$

$$\lambda_i \leq a^{\max}(1-e)$$

Finally, the maximization problem can be written in terms of actual abatement, a_i:

$$\underset{p_i, a_i}{Max} \left(1 - F_{(m)}^{-i}\right)\left(p_i a_i (1-e) - C_i\left(a_i\right)\right)$$

$$s.t. \qquad a_i \leq a^{\max}$$

Solving the nonpoint source profit maximization problem to determine optimal bidding strategy will give:

$$H = (1 - F_{(m)}^{-i})\left(p_i a_i (1-e) - C_i\left(a_i\right)\right) + k_1(a^{\max} - a_i).$$

This gives the following first order conditions:

[17] Weather uncertainty is incorporated into the choice of abatement strategy.

$$\frac{\partial H}{\partial p_i} \equiv -f_{(m)}^{-i}\left[p_i a_i(1-e) - C_i\left(a_i\right)\right] + (1-F_{(m)}^{-i})a_i(1-e) = 0$$

$$\frac{\partial H}{\partial a_i} \equiv (1-F_{(m)}^{-i})\left[p_i(1-e) - \frac{\partial C_i\left(a_i\right)}{\partial a_i}\right] - k_1 = 0$$

$$\begin{cases} k_1 > 0 \\ a_i = a^{,\max} \end{cases} \quad or \quad \begin{cases} k_1 = 0 \\ a_i < a^{\max} \end{cases}$$

Rewriting the first order condition taken with respect to p_i condition shows the optimal price bidding strategy for the surplus abatement strategy under the discriminating price auction: $p_i = \frac{C_i\left(a_i^*\right)}{a_i^*}\frac{1}{(1-e)} + \frac{(1-F_{(m)}^{-i})}{f_{(m)}^{-i}}$. Each bidder will inflate their bid above the reservation price . The term $\frac{(1-F_{(m)}^{-i})}{f_{(m)}^{-i}}$ is the typical hazard function commonly found in problems of adverse selection. The numerator is the probability of being selected into the team conditional on the bidder's reservation price, and the numerator is the change in the probability of being selected into the team corresponding to a unit increase in bid price above the reservation price. The greater the impact an increase in bid price has on the probability of team entry, the smaller the overall informational rent. Therefore, the nonpoint sources at the low-cost end of the distribution will extract greater informational rents than nonpoint sources at the high-cost end of the distribution. The optimal bid price line is always greater than the average abatement cost ensuring that the individual rationality constraint is always met given the Nash abatement strategy.

The remaining first order condition provides the optimal quantity bidding strategy. Rewriting the equation in terms of bid price, p_i, gives:

$$\frac{\partial H}{\partial a_i} \equiv (1-F_{(m)}^{-i})\left[p_i - \left(\frac{\partial C_i(a_i)}{\partial a_i}\frac{1}{(1-e)}\right)\right] - \frac{k_1}{(1-e)} = 0$$

By substituting the optimal bid price line for p_i into the equation, the optimal bid quantity can be represented in terms of the relationship between the optimal bid price line and the marginal cost of actual abatement (i.e., the marginal cost of abatement adjusted to account for the optimal abatement strategy).

$$\frac{\partial H}{\partial a_i} \equiv (1-F_{(m)}^{-i})\left[\left(\frac{C_i(a_i)}{a_i}\frac{1}{(1-e)} + \frac{(1-F_{(m)}^{-i})}{f_{(m)}^{-i}}\right) - \left(\frac{\partial C_i(a_i)}{\partial a_i}\frac{1}{(1-e)}\right)\right] - k_1 = 0$$

$$\begin{cases} k_1 > 0 \\ a_i = a^{\max} \end{cases} \quad or \quad \begin{cases} k_1 = 0 \\ a_i < a^{\max} \end{cases}$$

The slack conditions can be interpreted as follows. When the optimal bid price line is greater than the marginal cost of abatement at the maximum bid quantity, the nonpoint source bids the maximum. Otherwise, the nonpoint source will bid the expected profit maximizing quantity, where the optimal bid price line intersects the marginal cost of abatement curve (Figure 3.5).

Vickrey's expected revenue equivalency of the uniform and discriminating price auction, under risk neutrality assumptions, have been well documented (Harris and Raviv, 1981). Traditionally, this has made the seller's choice of auction design unimportant in terms of efficiency. However, in the case of the team entry auction,

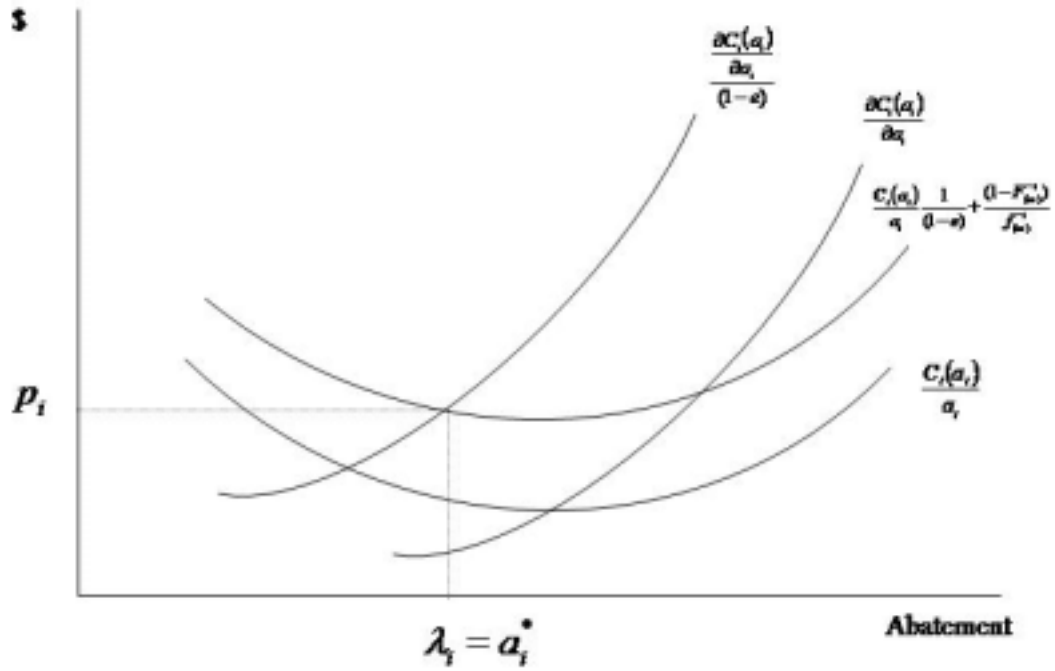

Figure 3.5: Optimal Bid Price and Quantity (Discriminating Price Auction and Surplus Abatement Strategy)

market efficiency can be greatly affected by the choice of auction design. Only the discriminating price auction can guarantee that the contract prices of all team members are consistent with the incentives necessary for the Nash abatement production strategy.

The stability of the Nash equilibrium abatement strategy is dependent on the provision of the optimal contract price for each team member. In the uniform price auction, the optimal bid quantity is based on an expected contract price, which may not be the same as the realized contract price. The only guarantee is that it will be greater than the winning bid prices. However, any contract price that differs from the expected price will result in each team member producing abatement at some quantity other than their bid quantity.

This problem does not arise in the discriminating price auction, where bid prices and quantities equal the contracted price and quantities for all winning bidders. When actual draws from the distribution of the mth order statistic differ from expectations, the effect is felt in the selection of team members. Bidders who thought they would be included in the team can be left out, or those expecting to be left out of the team can be selected into the team. In either case, there is no residual effect on the optimal provision of abatement from team members, since the bid price and quantity remain the contract bid and quantity for all nonpoint sources selected into the team, and it remains optimal for each to produce abatement equal to the bid level quantity.

3.8 Summary and Conclusions

The asymmetric information problems of inherent in contracting for nonpoint source pollution abatement can be overcome using a collective performance contract. The proposed two-stage contract pairs a team entry auction with an "all-or-nothing" budget breaking team contract. The auction overcomes adverse selection problems and provides crucial cost information to the point source to ensure that contracts meet participation and incentive compatibility constraints. The all-or-nothing team contract removes the free-riding incentive of moral hazard in teams. The contract provides a stable abatement production equilibrium.

The Nash equilibrium abatement strategy that emerges from the team contract is determined by the distribution of weather shocks. The point source must concede informational rents in order to ensure that the contracted level of abatement is produced, even without being able to observe individual emissions. The amount of informational rent is dependent upon the distribution of weather impacts on abatement in the watershed.

The use of a team entry auction will overcome the adverse selection problem in two ways. It allows the point source to sort potential team members in terms of relative abatement costs. The auction also provides the point source with contract prices and quantities that will meet the participation constraint and will ensure contract compliance. The choice of auction design is important. The team contract ensures that each nonpoint source will produce a contracted level of abatement only when the appropriate price is paid. In the uniform price auction the contract price can differ from the expected price

used by the nonpoint source to determine its optimal bid price and quantity. This can result in deviations from the contracted level of abatement being produced. The discriminating price auction guarantees the bid price and quantity when the collective target is reached. Because of this it is the preferred design for use in WQT trading markets.

CHAPTER 4

WHAT IS MORALLY SPECIAL ABOUT NONPOINT SOURCE POLLUTION?

This chapter examines the moral considerations that bear on the differential treatment of nonpoint sources of pollution with respect to the polluter-pays-principle, and addresses this broad question: *Is it unfair for US environmental policy to provide differential or special treatment to nonpoint sources?* The somewhat vague term *unfair* is used here in the more narrowly defined sense that "… an outcome (or process) is unfair if the treatment received by various individuals concerned fails to reflect the presence or absence of morally relevant differences between them" (Kagan, 1998 p. 54). I will argue that the point/nonpoint distinction is only morally relevant only under performance-based regulation, and even then does not provide a general moral prohibition against the application of the polluter-pays-principle to nonpoint sources.

First, the nature of pollution as an accumulative harm, and the resulting need for the state to redefine the harm threshold in terms of individual emissions standards are presented. The implication of regulatory control and the role of penalties as a morally justified means of enforcing compliance is contrasted with the notion of criminal

prohibition and the use of punishment. The moral relevance of the statutory distinction between point and nonpoint sources is discussed in terms of two commonly stated grounds: “differential uncertainty” and “differential observability”. I find there to be no moral relevance to the distinction based on differential uncertainty. Framing the distinction in terms of differential observability is morally relevant when considering performance-based regulations, but not technology-based regulations. In the performance-based case, the need for collective penalties results in the creation of vicarious liability. The moral legitimacy of extending the polluter-pays-principle to nonpoint sources will involve tradeoffs regarding the imposition of costs on innocent nonpoint sources under the polluter-pays principle, or imposing costs on the innocent public, under the exemption. Vicarious liability, in and of itself, is not morally objectionable. The conditions under which it is applied, as well as the degree of vicarious liability imposed, will play a crucial factor in the moral justification of its use. The chapter concludes with a discussion of some morally relevant factors that limit the moral objections to imposing vicarious liability under the polluter-pays-principle, tipping the scale in favor of its adoption for nonpoint source pollution.

4.1 Harm, Regulation, and the Polluter-pays-principle

The natural environment provides fundamental services without which humans, and most other life forms, could not survive. A few of these services are the provision of clean air and water, waste assimilation, climate control, and soil fertility. These

environmental services allow our society to produce its cornucopia of essential goods, including food, shelter, fuel, and medicines. The production of many of these goods results in the creation of unwanted byproducts, such as heat, chemicals, minerals, and nutrients. When emissions of these byproducts exceed the assimilative capacity of the natural environment, they impair ecosystem services, and are labeled pollution.

Pollution constitutes harm. In its simplest form, harm can be thought of as the production of something that is intrinsically bad. However, in many cases our actions do not directly produce something intrinsically bad, but instead alter the distribution of instrumental goods, which also can constitute doing harm (Kagan, 1998). Pollution is not only a harm when excess nitrogen in drinking water causes illness, but also when excess nitrogen indirectly deprives others of recreational or commercial uses of the water.

The harm from pollution can have an interesting structure, as emissions from one individual, in isolation, may be unnoticeable or innocuous. However, these emissions in combination with the emissions of others may cause serious harm. In this sense, pollution is often an accumulative harm[18], which presents the following challenges in terms of regulation:

> (i) A threshold of harm is approached, reached, or exceeded through the joint and successive contributions of numerous parties. (ii) These contributions are uneven in amount, and unequal in degree of care and social value. (iii) In respect to the harm of pollution, each contribution is "harmless" in itself except that it moves the condition of the environment to a point closer to the threshold of harm. (iv) When these accumulations cross the harm-threshold, they constitute public harms in that they set back vital net interests shared by almost everyone. (Feinberg, 1984 p. 225)

[18] There are many situations in which pollution is not an accumulative harm. For example, an oil tanker that spills its cargo in the waters off the Galapagos Islands crosses the harm-threshold on its own. However, within this chapter, I will be discussing pollution of the accumulative type, where no single point source or nonpoint source is producing enough emissions to cross the harm threshold alone.

Accumulative harms present difficulties in determining what should count as "doing harm" to others. The emissions from each source contribute to the production of harm, but do not cross the harm-threshold on their own. Therefore, there is no instantly recognizable basis for arguing that any one particular source has caused a harm, and hence to which particular source to impute liability for the harm. As a result, the state assigns sources a "permitted share" of the total allowable emissions, and only in reference to these individual emission standard can any particular polluting act be defined as wrongful.

Most of the activities that result in the production of unwanted emissions are socially desirable. For example, agriculture produces the food we eat, and power plants provide our electricity. An outright prohibition on the emission of unwanted byproducts would obviously cause more harm than the pollution itself[19]. For this reason, the role of the state is not to forbid polluting activities, but instead to regulate polluting activities. Hence, environmental statutes are the subject of administrative law as opposed to criminal law. The use of sanctions such as fines, enter only derivatively as a means of enforcing the regulatory authority of the state.

"If we are going to confer authority on designated officials in order to make some governmental program or institution work, we are committed thereby to granting them enforcement powers, since unenforceable authority is in effect, no authority at all" (Feinberg, 1984 p. 21). This arrangement is not unique to environmental regulation.

[19] Outright prohibitions are feasible in the case where the pollutant is extremely toxi or where the ban does not have serious economic impacts for society. Within this dissertation the focus has been on nutrients, such as phosphorous and nitrogen, which I stipulate are not extremely toxic or harmful and whose outright ban would have overwhelming negative impacts on society.

There are numerous statutes, that while not criminal statues, allow for the use of fines. For example, traffic laws that impose fines for operating a vehicle with a broken tail light or parking illegally.

The distinction between regulating and forbidding has important legal and philosophical implications, especially in terms of whether the criminal sanctions they may impose are considered “punishment” or “penalties”. Legally, the distinction between regulatory penalty and criminal punishment has constitutional implications. “There are elaborate constitutional safeguards for persons faced with the prospect of punishment; but these do not, or need not, apply when the threatened hard treatment merely regulates an activity” (Feinberg, 1965 p. 409). Penalties and punishments share the common feature of being state sanctioned and administered deprivations for transgressions, but punishment has an additional feature of expressing attitudes of indignation and reprobation of the state, the public, or both (Feinberg, 1965). The application of penalties, which do not contain this expressive function, is morally legitimate in many cases where punishment would not be. This is not meant to imply that penalties do not have to be morally justified, only that the justification does not involve the reference to any of the complex notions of condemnation, authoritative disavowal, symbolic non-acquiescence, vindication of the law, and absolution of others that are part and parcel of punishment (Feinberg, 1965).

The United States, along with a majority of the international community, has adopted the polluter-pays-principle (“PPP”) as the foundational tenet of environmental regulations as ratified in the Rio Declaration on the Environment, and the Organization

for Economic Cooperation and Development's Guiding Principles Concerning the International Aspects of Environmental Policy. The PPP requires that the costs of preventing and controlling pollution be borne by the polluter. Page has described it as a distributive principle, because it tells us in which direction cost bearing should be directed:

> The normative appeal of the principle rests in large part, I think, on a simple idea of fairness. Thus we say that it is fair for you to pay for the costs you impose on others, and more strongly, it is unfair for you to impose costs on others without bearing at least some cost. (Page, 1986 p.243)

Although a moral defense of the PPP is beyond the scope of this paper, the basis for its broad normative appeal can be briefly discussed. Kagan (1998) proposes categorizing normative theories into two classes, based upon the moral significance they place upon the outcome of an action. Consequentialist theories consider the goodness of an outcome the only morally significant factor in determining the rightness of an act, while plausible deontological theories consider the goodness of outcomes as only one of several, sometimes conflicting, morally significant factors. Consequentialists base the moral justification of a regulatory system in terms of the minimization of future harms, while the deontologist will traditionally appeal to a theory of moral desert. The quote from Talbot Page above captures the desert based appeal of the PPP. It directs the imposition of costs, abatement costs as well as any potential penalties for noncompliance, to those individuals who generate the costs. It targets the imposition of costs to those who "deserve" to bear them. Those who appeal to the moral desirability of the deterrent effects of regulatory control will also support the PPP.

As first shown by Coase (1960), given minimal transaction costs, the initial allocation of responsibility for cost bearing will not affect the ability of individuals to achieve the desired level of pollution reductions. Therefore, when these conditions hold the individual concerned solely with deterrence is indifferent regarding who should bear the costs. This indifference favors the choice of the PPP within a morally pluralistic society. This is because the PPP is morally unobjectionable based in terms of both desert and deterrence.

The PPP has two basic components. The first is compulsory compliance with state mandated standards. The second is that compliance is enforced through the use of penalties, such as fines. A highly stylized description of polluter-pays water quality protection regulations can be described as follows: (i) an ambient water quality standard is determined for an impaired waterway; (ii) specific standards are assigned to individual sources that discharge pollution into the waterway; (iii) the individual standards are set so that in aggregate, the ambient standard can be achieved (the standards can be technology-based or performance-based); (iv) individual sources are monitored, and noncompliant sources are punished through fines and penalties. The setting and enforcement of individual standards for pollution control can be a complicated process.

4.2 The Moral Relevance of the Point/Nonpoint Distinction

The Clean Water Act classifies pollution sources into two categories based upon the manner in which emissions are discharged. The Act defines a point source as “any

discernible, confined, and discrete conveyance" of pollutant to a water body (33 U.S.C. §1362(14)). A discrete conveyance includes, but is not limited to, "any pipe, ditch, channel, tunnel, conduit, well, discrete fissure, container, rolling stock, concentrated animal-feeding operation, landfill leachate collection system, vessel or other floating craft from which pollutants are or may be discharged" (33 U.S.C. §1362(14)). The definition of point source covers a wide and expanding variety of activities, beginning with direct discharges from factories and sewage treatment plants, but extending to a multitude of others.

The term nonpoint source is not defined directly by the Act, but instead is understood to be those pollution dischargers that are not covered within the definition of a point source. In addition, return flows from irrigated agriculture and agricultural storm water runoff are specifically excluded from the definition of a point source. This provides the majority of agricultural activities with nonpoint source status.

Nonpoint sources do not participate in the Clean Water Act's primary regulatory mechanism, the National Pollutant Discharge Elimination System. Instead, efforts to control agricultural nonpoint source emissions are dealt with through sections 208 and 319 of the CWA, which focus on state designed, taxpayer funded, subsidy-based programs in which participation is voluntary (33. U.S.C. §1329). Thus, the differential treatment of point and nonpoint sources of pollution is in terms of imputed responsibility and the application of the PPP. To be exempt from the PPP provides the nonpoint sources with the right to unrestricted emissions. The corollary to this is that any reductions cannot be required by the state, but instead must be made on voluntary

grounds, which, in the lion's share of cases, entails the use of public or private subsidies to offset the costs of abatement.

Many practical political, social, and economic explanations have been offered to explain the differential treatment of nonpoint sources, ranging from the strength of the agricultural political lobby to the historical social importance of agriculture (Ruhl, 2000). The focus of this chapter is whether a morally legitimate distinction between point and nonpoint sources can be made based on the statutorily defined partition of disperse and discrete conveyance. Within this more limited concern, two rationales are commonly given to explain why nonpoint sources may be exempted from the PPP. The first concerns differential uncertainty, which means that nonpoint source emissions are more variable and thus more difficult to control than point source emissions. The second is differential observability, which means that the inability to observe individual nonpoint source contributions prevents the morally legitimate application of the PPP.

4.2.1 Differential Uncertainty

Arguments in favor of the point/nonpoint source distinction being made in terms of differential uncertainty are asserting that the contribution to ambient conditions is inherently more uncertain in the case of nonpoint sources. Therefore, we should be reluctant to assign responsibility for violations of individual emissions standards. The empirical basis of this claim may be disputed, as both point source and nonpoint sources are stochastic. For example, the same wet weather event that can lead to an increased

runoff of nutrients from agricultural land, will also affect the emissions of nutrients from publicly owned wastewater treatment plants (POTWs). High precipitation levels will place greater demands on the POTWs and often results in excess nutrient in their effluent, in addition to the effects of increased output through combined sewer overflows (Huanxian, et al, 1997). The relative uncertainty of emissions between point and nonpoint sources is an empirical question that is site specific. Regardless, even if nonpoint source emissions are always more uncertain than point source emissions, in what way may "differential uncertainty" be morally relevant to the exemption of nonpoint sources from the PPP?

The appeal to differential uncertainty seems to be the following: If, in the face of uncertainty, an individual cannot be certain about the consequences of his actions, then he has no way of determining which action will produce the best results, and thus, which action is the right one. Quite simply, outcomes may not reflect intentions when uncertainty exists, and thus penalties for noncompliance may be undeserving. However, the consideration of intended actions is not required in this case. The administration of fines within environmental regulations is based on the violation of an emissions standard, irrespective of the intention of the source. "The rationale of strict liability in public welfare statutes is that violation of the public interest is more likely to be prevented by unconditional liability than by liability that can be defeated by some kind of excuse; even though liability without "fault" is severe, it is one of the known risks incurred by

businessmen" (Feinberg, 1965 p. 415)[20]. Uncertainty in terms of the appropriate action to take cannot absolve an individual from responsibility for the outcomes of those actions.

In fact, the argument for a distinction based on differential uncertainty, may in fact, work in the opposite direction. In many cases we feel justified in imposing stricter regulatory controls over those who choose to engage inherently risky activities. Consciously choosing to perform a risky act, while acknowledging having little control over the outcome, would appear to provide even greater moral legitimacy for the use of more stringent regulatory controls. This is seen with all types of ultra-hazardous activities to which we assign strict liability. Regardless, the appeal to differential uncertainty fails to provide a valid distinction for the exemption of nonpoint sources under the PPP.

4.2.2 Differential Observability

It is the ambient effect of pollution that concerns us in terms of harm prevention. However, the linkage between individual actions and their actual contribution to observed ambient effects is a complex and dynamic process of transport, spatial interactions, and stochastic natural phenomena. The costs of determining a reliable measure of causal link for all of the contributory individual actions that can come to bear on ambient effects are prohibitively high, if at all possible, to determine with certainty. Therefore, there is an

[20] This is one of the areas where the importance of the distinction between criminal "punishment" and regulatory "penalty" is quite evident. While we do not object to the practice of penalizing persons for "offenses" they did not mean to perform within public regulations such as issuing parking tickets, it is much different when considering doing the same in terms of punishing someone for the crime of murder (Feinberg 1965).

observability problem associated with ascertaining the contributory share of each individual source to the ambient water quality conditions. This problem is shared by both point and nonpoint sources. In order to control ambient water quality conditions, regulators must turn to a rougher approximation of contributions to ambient effects. Emissions are directly related to ambient conditions and pollution sources have immediate control over their production. As such, they have become the focus of pollution policy through two forms of emission standards: technology-based and performance-based.

4.3 Moral Relevance of "Differential Observability" Technology-based Regulation

Technology-based standards specify the abatement technology, technique, or practice that sources of potentially harmful emissions must use. That is, they impose control by dictating *how* emissions are to be reduced. Technology-based standards require the adoption of "best available control technologies", which provide an expected level of emissions reduction. The linkage to actual individual emissions reduction, and thus ambient effects, is indirect. However, this approach gives the regulator a reliable and enforceable basis for gaining control of emissions that contribute to those ambient conditions. "Doing harm" is necessarily redefined as failure to adopt the mandatory control technology, which is only indirectly related to the original concern of exceeding the accumulative harm-threshold. In this sense, adoption of technology indemnifies the source from responsibility for actual harm that may still occur. In the case of technology-

based standards, there is no relevant distinction between point and nonpoint sources. "Differential observability" does not exist in terms of the adoption of abatement technology. The installation of a nonpoint source filter strip is as easy to observe as the installation of a point source coal scrubber. In addition, both point sources and nonpoint sources can take (or fail to take) unobservable actions that reduce their overall abatement costs but also reduce the performance of the technology. Because of this, a morally relevant distinction between point and nonpoint sources cannot be made based on differential observability, when employing technology-based standards. Thus, the exemption of nonpoint sources from the PPP in terms of technology-based policy cannot be morally justified.

4.4 Moral Relevance of "Differential Observability" Performance-based Regulation

Limitations of the technology-based approach have the potential for abuse by point and nonpoint sources alike. Poorly funded or ineffective inspection regimes and the potential for shirking due to the presence of hidden action, as discussed in Chapter 2, will lead to ineffective water quality protection. This has led regulators to move towards greater use of performance-based instruments, which set emissions standards that cap the allowable level of emissions for each individual source. "Doing harm" is again redefined, this time in terms of exceeding one's individual emissions standard. Since

point source emissions are observable at the individual level, the application of the performance-based standard is unremarkable. However, this is where the distinction between point and nonpoint sources becomes relevant.

4.4.1 Performance-based Standard Setting

How unobservable are nonpoint source emissions? Unlike the point source, where there is an "end of pipe" from which actual emissions can be monitored, the dispersed nature of nonpoint sources defy direct measurement at a point of entry. This does not mean that the regulator is completely uninformed regarding individual contributions to collective nonpoint source emissions. Computer simulations that incorporate information on typical land use, soil type, slope, location, etc. can give realistic estimations of individual nonpoint source emissions (Griffin and Bromley, 1982). These estimations can then be used as a basis for allocating individual emissions standards. Basing individual standards on estimates of actual emissions will allow for some misallocation. However, the same can be said for the point source standards where individual emissions are observable. Setting a point source standard that accurately captures the actual contributions to resulting ambient conditions is improbable, because of the previously mentioned complex interactions that occur. Therefore, arguments can be made regarding the arbitrariness of the individual standards from either the point or the nonpoint source side.

In both the point source and nonpoint source case, a collective level of allowable emissions, which approximates the ambient damage threshold, is divided among the sources as a standard that specifies the individual allowable emissions level. The relative vagueness of the correlation between the assigned standards and the actual contribution to ambient damages has been previously acknowledged for both point and nonpoint sources. Rather than certainty, morality requires that the correlation between standards and actual contributions be as accurate as possible, considering the costs associated with further refinement and precision. The notion of wrongdoing is no longer predicated on actual ambient damage, but instead on violations of individual emissions standards. Therefore, once the emissions standards are assigned, any disconnect with the actual contribution to exceedances of the harm-threshold is inconsequential in the both the regulatory and normative sense. This leaves us without a moral distinction based on differential observability in terms of setting individual emissions standards.

4.4.2 Performance-based Standard Enforcement

It is in the enforcement stage that the use of an estimated emissions standard becomes truly troublesome, and in which case, the moral relevance of differential observability cannot be dismissed. Enforcement of individual standards on the point source side is relatively easy. Individual emissions can be observed and compared to the allowable level permitted under the standard. The assessment of fines and penalties can be directed at those point sources found to be noncompliant. Deviations from the

estimated individual emissions standard cannot be detected in the case of nonpoint sources. Violations of the aggregate nonpoint source emission standard can only be inferred from observations of violations of the ambient standards. The regulator is unable to determine which nonpoint sources are actually in violation of their individual standards. Since the noncompliant nonpoint sources cannot be identified, the regulator can only enforce the performance-based standard through the use of a collective-penalty.

4.5 Collective Penalties and Imposing Costs on the Innocent

Vicarious liability holds an individual liable for an outcome without a claim of direct causal contribution to the outcome. This is quite different from the concept of strict liability previously discussed. Under strict liability, an action need not be "faulty" to trigger an individual's responsibility for its outcome. In contrast, vicarious liability holds that it need not be an action performed by the individual that triggers his responsibility for an outcome. The distinction between punishments and penalties plays a role in the palatability of using strict liability within the enforcement mechanism. In the case of vicarious liability "… it does not follow that the safeguards of culpability requirements and due process which justice demands in the latter are always irrelevant encumbrances to the former" (Feinberg, 1965 p. 418). The moral legitimacy of imposing vicarious liability is dependent upon the conditions under which it is applied.

Vicarious liability can be categorized in two ways, vertical and horizontal. Vertical vicarious liability involves attaching responsibility for the outcome of an action

to individuals who are only related to the outcome through some formal hierarchal relationship, such as the principal-agent relationship[21] When an individual engages another to act as his representative, both parties can be held liable for the effects of the agent's independent actions. For example, the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) utilizes joint and several liability for the generation, hauling, and storage of hazardous waste.

The issue of including nonpoint sources in a collective penalty scheme is an example of horizontal vicarious liability. In this setting, there is no formal chain of control, or hierarchy, between individuals. This does not mean that it is inevitably unacceptable. Some examples of horizontal vicarious liability include voluntary suretyship and bonding. In such cases, an individual who neither authorizes, nor directs, nor carries out the actions of another, agrees to be responsible for potential outcomes of the other's action. For example, a parent may co-sign a car loan for a child, agreeing to be held vicariously responsible in the event of default. The voluntary acceptance of vicarious liability, especially in the horizontal formulation, is at least one factor that influences the palatability of its use.

When utilizing performance-based instruments for nonpoint source pollution, a collective penalty must be incorporated in order to avoid free-rider problems, which will occur within both subsidy and PPP policies. Collective penalties provide an incentive for each individual to be concerned with the performance of the collective as a whole. This

[21] The terminology "principal" and "agent" are being used here as in agency law. It is meant to denote a chain of command relationship or hierarchy, and the concern is not on how the action of the agent affects the principal directly (i.e., providing less than full effort), but rather how the action of the agent affects a third-party, and how this indirectly affects the principal. For example, if an agent is hired to deliver supplies and drives the principal's truck into a school bus, which is legally liable?

reduces the overall number of violations by removing the motivation to intentionally shirk from one's responsibility.

In Chapter 3, a collective performance-based subsidy approach was developed. Individual compliance in terms of the delivery of contracted levels of unobservable individual emissions reductions was enforced through a threat of collective penalty. If violations at the ambient level were detected individual payments were withheld for all members of the trading group. Thus, the collective penalty of withholding payment held all members of the trading group vicariously liable for failure to deliver the collective level of emissions reductions. A relevant consideration in the subsidy case, of course, is that participants join the abatement team voluntarily, with full knowledge of the vicarious liability dimensions of the contract. This is not the case under the PPP where participation is compulsory for all nonpoint sources. The application of a collective penalty (i.e., a collective fine) raises moral questions in terms of the potential for imposing costs on undeserving nonpoint sources. That is, using a collective fine penalizes noncompliant and compliant nonpoint sources alike.

Any morally legitimate water quality protection policy will aim to penalize all, but only, the sources that are guilty of violating their assigned standards. An injustice is done both when a non-offender is penalized, and when an offender is not. However, in reality, there is more often than not some degree of ambiguity associated with the determining who, in fact, was the offender. That is, the causal sequence of events that establish an individual's responsibility may be impossible to know with certainty, or the costs of such precision are prohibitively high. Therefore, within any regulatory

enforcement system, it is possible that an innocent individual may be penalized, or that a guilty individual may not. "In fact, in the realm of environmental issues, the role of empirical uncertainties is so pronounced that these uncertainties must seriously affect our thinking as to which direction political philosophy needs to take if environmental problems are to be addressed in a way that is both philosophically defensible and practically effective" (Ellis and Ravita, 1997 p. 209). When we lack sufficient information to avoid all rights violations, or injustices, we must retreat to trade-offs, on the grounds that any action, including the decision not to act, will yield some unjust results (Sher, 1984). To justify any regulatory system in which the rights of the innocents may be violated, it must be shown that these violations are justified by concerns of social good, and the rights of citizens to protection by the state (Philips, 1985). This means more than a simple weighing of the trade-offs of rights violations, but instead involves questions of the types of violations being imposed as well as their magnitude.

4.6 Justifying the Use of Collective-Penalties to Enforce the Polluter-Pays-Principle for Nonpoint Source Pollution

In addressing water quality issues through a performance-based approach, the regulator has only two options regarding nonpoint sources, both of which impose costs on innocent individuals. First, nonpoint sources can be regulated under the PPP, in which even the best-intended formulation of a collective-fine will penalize some nonpoint sources for the noncompliance of others. Second, the nonpoint sources can be exempted

from the PPP, which imposes the costs of forced contributions to subsidy payments on the innocent victims of pollution.

With the introduction of the differential observability distinction, the broad normative appeal of PPP, which was presented at the beginning of this chapter, is no longer obvious. Differential observability results in the regulator being unable to impose own-cost bearing, since the choice of either regulatory approach will impose costs on innocents. This suggests that the retributivist's preference for the PPP is no longer assured. In addition, differential observability introduces asymmetric information (i.e., the nonpoint sources know more about their own emission levels than the regulator), which means that the initial allocation of responsibility for cost bearing may affect the ability of individuals to achieve the desired level of pollution reductions with the same efficiency. This implies that the consequentialist may no longer be morally indifferent between the two regulatory approaches.

Given this lack of obvious consensus between normative theories, a promising approach for determining the morally legitimate regulatory policy is to take a broader foundational stance by combining considerations of both desert and deterrence within a consequentialist theory. Although consequentialists define moral permissibility of acts only in terms of the relative goodness of their outcomes, they can still incorporate distributional concerns within their definition of what is good. For example, a consequentialist theory can be sensitive to the distribution of costs based on individual desert. This remains a purely consequentialist theory, as the goodness of the outcome, compared to all alternative actions, is the only factor in determining rightness. However,

goodness is judged, at least in part, by reference to the distribution of costs, in particular that individuals should bear their own costs.

4.6.1 A Sketch of a PPP Collective Performance-based Program

Under the PPP, the regulator sets individual emission standards for each nonpoint source. Allocated emission standards will sum to the total allowable level of collective nonpoint source emissions in the waterway. A collective fine will be imposed on all nonpoint sources when total emissions exceed the total allowable level, and the fine increases with the magnitude of the violation. The collective fine does not distinguish between compliant and noncompliant sources. Thus, during periods of noncompliance, some nonpoint sources have additional costs imposed on them for the action of others. The collective fine must be set high enough to remove the economic incentive to free-ride. In this setting, nonpoint sources bear their own costs of abatement required to meet their individual emission standard. Therefore, violations of individual emissions standards are minimized to the impacts of random events (e.g., weather effects and technology failures), human error, and perhaps, malicious acts.

4.6.2 A Sketch of a Collective Subsidy-based Program

When nonpoint sources are exempted from the PPP, a voluntary collective subsidy approach will be adopted in its place. In this case, the emissions of nonpoint sources are not restricted, and pollution is allowed to befall the innocent public. The only

option for reducing the nonpoint source pollution is to divert public funds to pay the nonpoint sources to voluntarily reduce emissions. The same total allowable level of emissions as under the PPP is desired by society. The regulator announces a per unit price that will be paid for nonpoint source emission reductions. Nonpoint sources that are interested in providing emission reductions submit bids for the quantity of promised abatement that they are willing to produce at this price. Nonpoint sources are admitted into the contract team in order of their bid quantities. The nonpoint sources that bid the largest quantities of abatement are selected first, until the socially desired level of abatement is reached. The regulator must set the subsidy price high enough to attract enough bids to achieve the socially desired level of total nonpoint source abatement, which, even in the case of very high subsidy payments, may not be possible. The collective-penalty in this case is the withholding of all subsidy payments when the team target is not achieved. This mechanism effectively removes the hidden-action incentive to free-ride. Each nonpoint source expends the costs to produce its contracted level of abatement. When the team target is achieved, each nonpoint source receives a subsidy payment based on its individual bid quantity of abatement and the contract price. Thus, the nonpoint source costs of abatement are imposed on the public only when compliance is achieved. When the team target is not met, individual nonpoint source subsidy payments are withheld, and each nonpoint source bears its own abatement costs.

4.6.3 Consequentialist Concerns

The consequentialist concern is with the minimization of future violations which is often framed in terms of deterrence. Within a single contracting period, both the PPP and collective subsidy programs provide the same level of protection against intentional, economically motivated deviations from individual abatement requirements. However, this still leaves the potential for violations attributable to random events, human error, and malicious acts. The consequentialist prefers the mechanism that can provide for the minimization of future violations related to these other causes as well. In the case of random events such as weather effects and technology failure, the collective penalty schemes of both programs encourage the adoption of abatement strategies that, in the presence of relatively high uncertainty, encourage self-insurance by producing more than the required level of abatement. However, the latter two causes of violations, human error and malicious acts, require a more cooperative solution that may best be served in the compulsory setting. Under both forms of collective-penalty, peer pressure can emerge as a self-policing tool within the group, and its efficiency in collective production enterprises has been well documented (Kandel and Lazear, 1992; Barron and Paulson Gjerde, 1997). The use of vicarious liability as a self-policing tool, where practical policing is otherwise infeasible, has been argued as a moral justification for its application (Feinberg, 1965 p. 681).

However, the communal ties and social capital required for effective peer pressure are likely to become well developed in the compulsory program, where a nonpoint source cannot escape the accident prone or malicious individual, and the peer group remains

largely unchanged from period to period. In the voluntary arrangement, the same continuity may not develop, and the ability for individuals to opt out of future participation would work against the incentive to invest heavily in effective peer monitoring and peer pressure institutions.

The frequency of harm is likely to be greater for the public under the collective subsidy approach than the harm imposed on the nonpoint sources under the PPP. This is because using either an efficient collective fine or collective subsidy, the ambient water quality is expected to be met more often than not. Thus, the public would bear the costs of the nonpoint sources (paying for emissions reductions) more often than the compliant nonpoint sources would bear the costs of noncompliant nonpoint sources (paying the fine for noncompliance). We are left having to decide between two principles that both implicitly authorize the imposition of costs on undeserving individuals. Within the PPP, all of the compliant nonpoint sources are penalized along with all of the noncompliant sources. However, this happens only when violations occur, which are reduced by the presence of the collective penalty. Under the subsidy based approach, the compliant nonpoint sources that may be mistreated as being noncompliant have agreed to such treatment in advance, thus removing the moral sting of the collective-penalty. However, pollution is still allowed to occur, and its harm negatively affects the general public, unless public funds are diverted from other beneficial uses to subsidize its removal. In addition, the required public payments will occur with greater frequency than the imposition of collective fines on nonpoint sources.

4.6.4 Retributivist Concerns

The retributivist concern is that costs should be borne by the offender, both the pollution generator and the noncompliant point source, and not imposed upon the innocent. As can be seen in the sketches provided above, the retributivist is most concerned with the imposition of costs in the PPP approach during periods of nocompliance, and the imposition of costs in the collective subsidy approach during periods of compliance. At all other times, both policy regimes are operating in accordance with the own-cost bearing principle.

The plight of the innocent nonpoint sources under the PPP differs from that of the public under the collective subsidy in two fundamental ways. Unlike the nonpoint sources, the public does not directly contribute to the accumulation of emissions that cause pollution. Therefore, they have no connection to the production of the pollution, while even compliant nonpoint sources are contributing emissions, albeit at state permitted levels, to the accumulative harm. "It may not be legitimate to require taxpayers to support government expenditures (either directly or indirectly) for beneficial activities with respect to which the people collectively do not have a duty or responsibility to take action" (Ellis and Ravita, 1997 p. 223).

The imposition of vicarious liability under the PPP is targeted to the group of individuals that are known to be contributing to the accumulation of pollution. In addition, their choice of land use is undertaken voluntarily and with prior knowledge of the potential for adverse side effects. Thus, the imposition of vicarious liability through

regulatory control is not comparable to a situation of holding innocent bystanders accountable for unanticipated outcomes. The individual performance of any nonpoint source is dependent upon three factors: abatement technology adopted, the underlying heterogeneous site and resource characteristics, and the nonpoint source's effort level. As has been argued, the regulator has the ability to observe the first two factors. It is only the effort level which remains impractical to observe. Thus, any vicarious liability is reduced to problems of misidentification of actual effort levels. While this does not eliminate the concern with imposing vicarious liability, it does greatly reduce the scope of the moral objection of applying the PPP to nonpoint sources.

4.7 Some Practical Factors That Should Be Considered

The previous section does not provide a universal determination regarding the moral legitimacy of the PPP in terms of nonpoint source pollution regulation. The weighing of relative harms depends upon the unique conditions of the regulatory situation being examined. However, there are some morally relevant factors that support the justification of the PPP in all situations.

An important consideration is the level of information held by the nonpoint sources. In order for the PPP to be morally legitimate, the nonpoint sources must at least be aware that they are contributors to an accumulative harm. Additional levels of knowledge regarding their own emissions, the performance of their abatement practices,

and that of their neighbors increases the argument in favor of the PPP. These information requirements are fostered by collectives that consist of a relatively small group of geographically proximate nonpoint sources. Throughout this chapter, the term nonpoint source has been used as a single category of polluters. However, the types of land use that can fall under the nonpoint source category are quite diverse. Urban stormwater runoff from cityscapes and suburban development, municipal and private golf courses, construction sites, and livestock and crop farming are just some of the land uses that fall under the nonpoint source umbrella. In addition, the information needed to estimate joint contributions to the aggregate level of pollution will be facilitated through a shared understanding of the activities generating the emissions.

Shifting the burden of proof, including act of God exemptions, and providing equal treatment are three due process considerations that can lend support to the moral justification of the PPP for nonpoint source pollution control. The PPP regulatory policy can allow individual nonpoint sources to provide evidence of compliance to avoid the collective fine. This capitalizes on the available information regarding adopted technologies and practices to provide some basis for distinguishing between the "deserving" and "undeserving" nonpoint sources. The amount of "evidence" that needs to be produced to provide immunity determines the coarseness of the collective fine. Act of God provisions exempt the nonpoint sources from collective fines when violations of the aggregate standard can be attributed to extreme weather events such as tornados or floods. This ensures that the nonpoint sources are not penalized for unforeseeable or uncontrollable conditions. Finally, providing equal treatment to similar nonpoint sources

within the collective is essential. This can be accomplished by ensuring that the fines imposed on the nonpoint sources reflect their probable proportional share to the aggregate emissions. For example, if slope is a primary factor in emissions, then nonpoint sources should be classified in terms of steepness, and individuals within those classes should get identical treatment.

4.8 Summary and Conclusions

Distinguishing between point and nonpoint sources of emission based on differences in the ability to observe individual emission levels is a morally relevant factor in the regulatory decision regarding the application of the PPP to nonpoint source pollution control. This moral relevance is limited to the case of performance-based policy instruments, as it is the enforcement of such policies that introduces the need for a collective penalty, and the issues of vicarious liability. Performance-based regulation under the PPP requires that the regulator make unavoidable trade-offs in imposing costs on the innocent. When nonpoint sources are exempted from the PPP, the regulator requires the innocent victims of pollution (i.e., the public) to divert valuable public funds to compensate nonpoint sources for voluntary emissions reductions. Under the PPP, the inability to observe individual abatement effort levels results in the inability to target only the noncompliant through the collective fine.

The legitimacy of the tradeoff between the imposition of costs on the innocent that is associated with the PPP and the collective subsidy scheme will depend upon a

myriad of factors that include, but are not limited to, the impact of pollution, the relative magnitudes of these costs, the frequency of their imposition, and the manner and environment in which the costs are imposed. No general moral barrier exists which prohibits the use of the PPP in relation to nonpoint source pollution. Instead, the moral legitimacy of its application will depend upon the unique characteristics of the situation in which it is being applied.

CHAPTER 5

CONCLUSIONS

The inability to directly observe the emissions of individual nonpoint sources does distinguish them from point sources. This makes the direct application of conventional regulatory instruments difficult. During the past thirty years, the Clean Water Act has been successful in reducing the negative impact of point source pollution. However, as this dissertation has shown, the water quality problems associated with nonpoint source pollution have proven to be more difficult to address, and currently pose the greatest threat to the nation's waters. As a result, the United States Environmental Protection Agency has begun to encourage innovative, incentive-based instruments to address nonpoint source water pollution.

Trading programs promise a means for achieving water quality standards at a lower overall cost. The traditional efficiency gains, attributable to the flexibility provided by trading, are coupled with the inclusion of nonpoint sources as a potential source of relatively low-cost abatement. In order to realize these potential benefits, two fundamental questions must be answered: (i) How will trades between point and nonpoint

sources be monitored and enforced?; and, (ii) How will nonpoint sources be included within a trading market?.

Trading between point and nonpoint sources can be accommodated in one of two ways. The first is a technology-based approach that allows for monitoring and enforcement of trades at the individual level. The second is a performance-based approach that requires monitoring and enforcement at the collective level.

The technology-based approach accommodates trading through the use of a proxy for unobservable, individual nonpoint source emission reductions. Trades are monitored and enforced based on the expected emission reductions provided from the adoption of specific abatement technologies. Uncertainty regarding the relationship between expected and actual abatement technology performance can be dealt with through the introduction of trading ratios.

The trading ratio specifies the number of units of expected nonpoint source pollution reduction that must be exchanged for a single unit increase in point source pollution. The optimal trading ratio depends on the expected performance of nonpoint source abatement technology, as well as the uncertainty associated with nonpoint source emissions. In practice, the optimal trading ratio will be set at a number greater than one. One reason is that the relative uncertainty of point and nonpoint emissions and abatement technology performance is *a priori* unknown, and the regulator is likely to act conservatively in setting the trading ratio. In addition, water quality trading markets are most often being adopted in areas where current ambient standards are not being met. In this case, a trading ratio greater than one is required to overcome the existing

noncompliance within the market. While a trading ratio greater than one ensures that the uncertainty associated with the proxy for actual nonpoint source emissions does not result in violations of the ambient standards, it also reduce the benefits of exchange by making nonpoint source offsets more expensive.

Trading ratios can effectively deal with the uncertainty associated with a technology-based approach, but they cannot deal with the more troublesome issue of hidden action. When actions that affect the performance of abatement technology are costly and unobservable, the potential for shirking arises. In the process of reducing the cost of abatement, the individual nonpoint source also reduces the effectiveness of the abatement technology. This requires the regulator to employ even larger trading ratios. At a minimum, the presence of hidden action is a source of market inefficiency. More problematic is the potential that the hidden action adjustment could lead to prohibitively high trading ratios, removing all benefits of trading and resulting in market collapse.

An alternative to the technology-based approach is to monitor and enforce trades on the basis of performance. The promise of this approach is that it removes the regulator's uncertainty about the effectiveness of the nonpoint source abatement technologies. This eliminates the need for a trading ratio, which provides an increase in the potential of trading. Uncertainty as to the effectiveness of adopted abatement technology is borne by the nonpoint sources, who are best able to handle it, allowing greater efficiency in the permit market.

However, as has been conceded, monitoring individual nonpoint sources on the basis of performance is technically difficult, and thus likely to be prohibitively expensive.

Therefore a performance-based approach must accommodate trading based on collective monitoring at the sub-watershed level. The difficulty of this approach is not in the monitoring of ambient water quality, which can be done with relative ease, the challenge comes in providing the right incentives to individual nonpoint sources to link the collective performance to appropriate individual actions.

Collective performance-based trading requires the point source to simultaneously contract with multiple nonpoint sources for reductions of individual emissions, while being unable to observe individual productivity. This problem is often referred to as "moral hazard in teams", and is similar to the problem of free-riding. If an individual does not fulfill his obligation to reduce emissions, this breach of contract is only detected at the collective level. In this way, the effect of individual shirking has the potential to spread across all agents in the group, as it cannot be attributed to the responsible party or parties.

Even though the actions of the nonpoint sources are not observable, and so cannot be used as the basis of the contract, the sum of the individual actions is verifiable as collective abatement. Therefore, to overcome the free-rider problem, the collective abatement outcome must be included in the contract that stipulates payment to the nonpoint sources. A successful contract must pay more when the observable collective performance is a good signal that the individual abatement choices were the required ones. The contract offered by the principal must make each agent feel responsible for the whole of the final product, in order to provide the appropriate incentive for overcoming the free-rider problem.

A collective performance based contract that pairs a team entry auction and an "all-or-nothing" budget breaking team contract can overcome the asymmetric information problems facing the point source, and accommodate performance based trading. The point source can determine which nonpoint sources to trade with through the use of an auction. The team entry auction requires the individual nonpoint sources to compete for selection into the abatement production team by bidding to reduce a specific amount of pollution at a particular price. Nonpoint sources participate in the market as profit-seekers, but the auction creates the incentive for nonpoint sources to limit rent seeking by keeping bid prices low. Only those nonpoint sources with the lowest bid prices are offered contracts. In addition, the bid information can then be used to formulate contracts that are known to be acceptable to the nonpoint sources.

Overcoming the moral hazard in teams problem is accomplished through the use of an "all-or-nothing" collective penalty for noncompliance. Individual contract payments are made only when the observed level of collective abatement meets or exceeds the sum of the individually bid quantities of abatement. This mechanism removes the profitability of free-riding for every member of the group, and leads to an enforceable contract. The nonpoint sources bear the uncertainty associated with weather shocks and technology failure, and will adopt abatement strategies that reflect these risks. The cost of these strategies is passed on to the point source in the form of additional rents.

Therefore, the exchange of point and nonpoint source emissions can be accommodated in a trading market in either a technology-based or performance-based

approach. While the technology-based approach allows for a more direct basis for individual monitoring and enforcement of trades, it is susceptible to problems of shirking. The use of a trading ratio to deal with the uncertainty of the proxy, cannot deal with the problem of hidden action without market inefficiency and potential market collapse. The alternative use of performance-based trading approaches requires the use of team contracts that provide individual incentives linked to the performance of the entire group. Such contracts must be designed to overcome both adverse selection and moral hazard problems. Performance-based approaches promise efficiency gains in terms of reducing the problems of asymmetric information, and by introducing flexibility into the choice of nonpoint source abatement technologies and practices.

Traditional tradable permit markets institute a cap-and-trade approach. That is, the total emissions for the market are capped at an ambient level and then trade is introduced. This requires all emission sources to hold individual emission standards, providing better ambient pollution protection. For water quality trading this would require nonpoint sources to be brought into the permitting process of the Clean Water Act (i.e., issuing and enforcing nonpoint source emissions standards for all nonpoint sources).

This is problematic as nonpoint sources have had a long standing tradition of exemption from direct regulation under the polluter-pays-principle. As a result, nonpoint sources are included in trading markets only on a voluntary basis, and they participate in return for compensation. In order to adopt a baseline cap for nonpoint source emissions we would have to be justified in applying the polluter-pays-principle to nonpoint sources.

But is there something morally special about nonpoint sources of pollution that prevents this?

Distinguishing between point and nonpoint sources of emission based on differences in the ability to observe individual emission levels is a morally relevant factor in the regulatory decision regarding applying the polluter-pays-principle to nonpoint source pollution control. This moral relevance is limited to the case of performance-based policy instruments, as it is the enforcement of such policies that introduces the need for a collective penalty, and the issues of vicarious liability. Performance-based regulation requires the regulator to make unavoidable trade-offs in imposing costs on the innocent. Under the polluter-pays-principle, a collective fine will impose costs on both compliant and noncompliant nonpoint sources alike. On the other hand, exempting nonpoint sources from the polluter-pays-principle requires the innocent victims of pollution (i.e., the public) to divert valuable public funds to compensate nonpoint sources for voluntary emission reductions. Both of these options are morally objectionable, but the regulator must choose. The morally justified choice will depend upon the unique social, economic, and practical conditions underlying each specific case. However, there is no general moral barrier to prohibit the application of the polluter-pays-principle to nonpoint sources of pollution. In this case, adopting more protective cap-and-trade programs should be possible in some water quality trading markets.

Limitations and Extensions

The research in this dissertation supports the development of collective performance-based water quality trading markets. The theoretical model predicts that the group contract will provide the appropriate incentives to achieve a stable equilibrium within water quality trading markets. This result takes advantage of the concept of implementing these markets in relatively small geographic regions. In this setting ambient water quality conditions provide an accurate mapping to collective abatement performance. In addition, concentrating the market in areas of local contiguous land parcels increases the information level that market participants possess regarding the performance of others to the collective goal, and to market efficiency.

When considering the actual implementation of these contracts, however, some practical concerns regarding the use of small localized markets must be addressed. In particular, the fact that market participants will have strong neighborhood and personal histories that can influence market interactions. These existing social bonds could conflict with the rationality assumptions underlying the economic predictions, leading to unexpected results in practice. Social science research that examines the role of social capital and existing social relations on individual behavior within collective performance-based water quality trading markets is required to gain a better perspective upon practical implementation.

As is the case of all mechanism design research, implementation of the proposed market contract requires us to address the fact that, in practice, individuals may not

behave in accordance with our theoretical axioms of rationality. For example, it has been shown that, unlike the assumptions of expected utility theory, individuals tend to frame decisions regarding equal probable losses and equal probable gains asymmetrically. Experimental research into practical decision-making responses to the incentives provided through collective-performance based contracts will move us towards more realistic contract designs.

Finally, the contract that is presented in this model is represented as a "one-shot" game. In practice, these contracts will be repeated over a longer time frame. Whether repetition of the team contract will result in refinement and stability of the market, or in market collapse cannot be predicted from the static model. A dynamic model that incorporates the complicated learning effects from observations of group performance would be required to gain a better understanding of the dynamic strategies of market participants.

Summary

Collective performance-based trading has advantages over the currently employed technology-based approach. First, it shifts the risk of noncompliance to the nonpoint sources that are better able to deal with it. Second, collective performance-based trading provides flexibility in terms of nonpoint source abatement technology choices. This provides the impetus for increased dynamic efficiency gains in terms of induced innovation effects. Problems of adverse selection and moral hazard in teams can be

overcome through the development of appropriate contract incentives, and the threat of collective penalties. Finally, the use of collective performance-based trading under the PPP can be morally justified allowing for the development of cap-and-trade markets for water quality trading.

APPENDIX A

Variable	***Definition***
ω	*Probability of good weather*
$1\text{-}\omega$	*Probability of bad weather*
$e:(0 \le e \le 1)$	*Weather shock*
$1+e$	*Impact of good weather on NPS abatement*
$1\text{-}e$	*Impact of bad weather on NPS abatement*
n	*Total number of NPS in watershed*
$m:(m \le n)$	*Total number of NPS in team contract*
λ_i	*Quantity of NPS abatement production bid by agent i*
p_i	*Price per unit of NPS abatement bid by agent i*
$\Lambda = \sum_{i=1}^{m} \lambda_i$	*Total quantity of NPS abatement contracted from team*
a_i	*Quantity of abatement produced by NPS i*
$A(a,e): \begin{cases} (1+e)\sum_{i=1}^{m} a_i \\ (1-e)\sum_{i=1}^{m} a_i \end{cases}$	*Quantity of NPS abatement observed in good weather and bad.*
$C_i(a_i)$	*Agent i's NPS abatement cost function*

Table A.1: Definitions of Notation Used in Chapter 3

APPENDIX B

SHORTFALL ABATEMENT STRATEGY

Assume that the point source sets the optimal sharing rules (p_i , λ_i) such that the participation constraint of each team member will be met, $p_i \geq \frac{C_i(a_i^*)}{\lambda_i}$. average costs. With these assumptions in place, it will be shown that the abatement production strategy that maximizes expected profit, contingent upon meeting the target in both good and bad weather, is $a_i = \frac{\lambda_i}{(1+e)}$. To show this, let $a_j = \frac{\lambda_j}{(1+e)}$ be the abatement production strategy for all nonpoint sources $j \neq i$, who were selected into the team. Nonpoint source *i*'s profit maximizing choice of abatement is determined by maximizing expected utility:

$$\underset{a_i}{Max} \qquad (\alpha) + (\beta)$$

$$where \quad (\alpha) \equiv \omega\left[\left(p_i\,\lambda_i\right) I\left((1+e)\left(\left(\frac{1}{(1+e)}\sum_{j\neq i}\lambda_j\right) + a_i\right) \geq \Lambda\right) - C_i\left(a_i\right)\right] \quad .$$

$$and \quad (\beta) \equiv (1-\omega)\left[\left(p_i\,\lambda_i\right) I\left((1-e)\left(\left(\frac{1}{(1+e)}\sum_{j\neq i}\lambda_j\right) + a_i\right) \geq \Lambda\right) - C_i\left(a_i\right)\right]$$

It has been assumed that the nonpoint sources are choosing the expected profit maximizing abatement strategy that will guarantee payment in good weather, but not in

bad weather. This is equivalent to choosing the expected profit maximizing level of abatement that guarantees $I(\cdot)=1$ in *(α)* and $I(\cdot)=0$ in *(β)*.

Solving both indicator functions from provides the minimal abatement levels at which the collective target will be met in each weather period, given the assumption regarding the abatement of the other nonpoint sources. So the indicator function from part (α) is solved as follows:

$$(1+e)\left(\left(\frac{1}{(1+e)}\sum_{j\neq i}\lambda_j\right)+a_i\right)\geq\Lambda$$
$$=\left(\frac{(1+e)}{(1+e)}\sum_{j\neq i}\lambda_j\right)+(1+e)a_i\geq\left(\sum_{j\neq i}\lambda_j\right)+\lambda_i$$
$$=a_i\geq\frac{\lambda_i}{(1+e)}$$

Prior to solving the indicator function from part (β), it must be modified as follows:

$$(1-e)\left(\left(\frac{1}{(1-e)}\sum_{j\neq i}\lambda_j\right)+a_i\right)<\Lambda$$

This is because we are interested in all abatement levels at which the indicator function will not hold.

$$(1-e)\left(\left(\frac{1}{(1+e)}\sum_{j\neq i}\lambda_j\right)+a_i\right)<\Lambda$$
$$=\left(\frac{(1-e)}{(1+e)}\sum_{j\neq i}\lambda_j\right)+(1-e)a_i<\left(\sum_{j\neq i}\lambda_j\right)+\lambda_i$$
$$=(1-e)a_i<\left(\sum_{j\neq i}\lambda_j\right)+\lambda_i-\left(\frac{(1-e)}{(1+e)}\sum_{j\neq i}\lambda_j\right)$$
$$=a_i<\sum_{j\neq i}\lambda_j\left(\frac{1}{(1-e)}-\frac{1}{(1+e)}\right)+\frac{\lambda_i}{(1-e)}$$

In order for $a_i \geq \frac{\lambda_i}{(1+e)}$ to be the optimal abatement strategy it must be true that $\frac{\lambda_i}{(1+e)}$ is always less than $\sum_{j \neq i} \lambda_j \left(\frac{1}{(1-e)} - \frac{1}{(1+e)} \right) + \frac{\lambda_i}{(1-e)}$. This will ensure that the stated condition that the collective target is only met in good weather will hold.

$$\frac{\lambda_i}{(1+e)} < \sum_{j \neq i} \lambda_j \left(\frac{1}{(1-e)} - \frac{1}{(1+e)} \right) + \frac{\lambda_i}{(1-e)}$$

$$\lambda_i < \sum_{j \neq i} \lambda_j \left(\frac{1}{(1-e)} - \frac{1}{(1+e)} \right)(1+e) + \frac{(1+e)}{(1-e)} \lambda_i$$

$$\Lambda - \sum_{j \neq i} \lambda_j < \frac{(1+e)}{(1-e)} \left(\Lambda - \sum_{j \neq i} \lambda_j \right)$$

Since the r.h.s. of the equation is always greater than the l.h.s., $a_i \geq \frac{\lambda_i}{(1+e)}$ is the optimal abatement strategy.

Finally, it must be shown that the abatement strategy actually holds at equality, $a_i = \frac{\lambda_i}{(1+e)}$, by maximizing expected profits, subject to the chosen abatement strategy.

$$\underset{a_i}{Max} \quad \omega\left(p_i \lambda_i (1) - C_i(a_i)\right) + (1-\omega)\left(p_i \lambda_i (1) - C_i(a_i)\right)$$

$$s.t. \quad a_i \geq \frac{\lambda_i}{(1+e)}$$

Construct the Hamiltonian:

$$\underset{a_i}{Max} \quad H = p_i \lambda_i - C_i(a_i) + k_1 \left(a_i - \frac{\lambda_i}{(1+e)} \right)$$

$$\frac{\partial H}{\partial a_i} \equiv \frac{\partial C_i(a_i)}{\partial a_i} = k_1$$

$$\begin{cases} k_1 > 0 \\ a_i = \dfrac{\lambda_i}{(1+e)} \end{cases} \qquad \begin{cases} k_1 = 0 \\ a_i > \dfrac{\lambda_i}{(1+e)} \end{cases}$$

The slack condition, $k_1 > 0$, holds with inequality, since $\frac{\partial C_i(a_i)}{\partial a_i} > 0$. Therefore, the abatement strategy holds with equality, $a_i^* = \frac{\lambda_i}{(1+e)}$. The rationale being that at equality the abatement production strategy ensures payment in both good and bad weather. Any additional abatement will increase costs but will not change the expected revenue.

APPENDIX C

NONPARTICIPATING STRATEGY

Assume that the point source sets the optimal sharing rules (p_i , λ_i) such that the participation constraint of each team member will be met, $p_i \geq \frac{C_i(a_i^*)}{\lambda_i}$. average costs. With these assumptions in place, it will be shown that the abatement production strategy that maximizes expected profit, contingent upon meeting the target in both good and bad weather, is $a_i = 0$. To show this, let $a_j = 0$ be the abatement production strategy for all nonpoint sources $j \neq i$, who were selected into the team. Nonpoint source *i*'s profit maximizing choice of abatement is determined by maximizing expected utility:

$$\begin{aligned} &\underset{a_i}{Max} \qquad (\alpha) + (\beta) \\ &where \quad (\alpha) \equiv \omega\left[\left(p_i \lambda_i\right) I\left(\left(1+e\right)\left(a_i\right) \geq \Lambda\right) - C_i\left(a_i\right)\right] \quad . \\ &and \quad (\beta) \equiv (1-\omega)\left[\left(p_i \lambda_i\right) I\left(\left(1-e\right)\left(a_i\right) \geq \Lambda\right) - C_i\left(a_i\right)\right] \end{aligned}$$

It has been assumed that the nonpoint sources are choosing the expected profit maximizing abatement strategy that will guarantee that no payment is received in either good weather or bad weather. This is equivalent to choosing the expected profit maximizing level of abatement that guarantees $I(\cdot) = 0$ in both *(α)* and *(β)*.

Solving both indicator functions from provides the abatement levels at which the collective target will be met in each weather period, given the assumption regarding the abatement of the other nonpoint sources. Prior to solving the indicator function from part (α), we modified it such that: $(1+e)(a_i) < \Lambda$, because we are interested in the abatement levels that will ensure that the condition is not met.

$$(1+e)(a_i) < \Lambda$$
$$= a_i < \frac{\Lambda}{(1+e)}$$

Similarly, the indicator function from part (β), must be modified as: $(1-e)(a_i) < \Lambda$

$$(1-e)(a_i) < \Lambda$$
$$= a_i < \frac{\Lambda}{(1-e)}$$

The abatement strategy that will ensure that the target is not regardless of weather is the strategy that produces less abatement. It is obvious that $a_i < \frac{\Lambda}{(1-e)}$ will produce less abatement than $a_i < \frac{\Lambda}{(1+e)}$, and is therefore the optimal strategy.

Abatement is constrained to be non-negative, thus is bounded by zero. Determining whether the strategy holds at zero is accomlplished by maximizing expected profits, subject to the chosen abatement strategy.

$$\underset{a_i}{Max} \quad \omega\left(p_i\,\lambda_i\,(1) - C_i\,(a_i)\right) + (1-\omega)\left(p_i\,\lambda_i\,(1) - C_i\,(a_i)\right)$$
$$s.t. \quad a_i \geq 0$$

Construct the Hamiltonian:

$$\underset{a_i}{Max} \qquad H = p_i \lambda_i - C_i(a_i) + k_1\left(a_i - \frac{\lambda_i}{(1+e)}\right)$$

$$\frac{\partial H}{\partial a_i} \equiv \frac{\partial C_i(a_i)}{\partial a_i} = k_1$$

$$\begin{cases} k_1 > 0 \\ a_i = 0 \end{cases} \qquad \begin{cases} k_1 = 0 \\ a_i > 0 \end{cases}$$

The slack condition, $k_1 > 0$, holds with inequality, since $\frac{\partial C_i(a_i)}{\partial a_i} > 0$. Therefore, the abatement strategy holds with equality, $a_i^* = \frac{\lambda_i}{(1+e)}$. The rationale being that at equality the abatement production strategy ensures payment in both good and bad weather. Any additional abatement will increase costs but will not change the expected revenue.

APPENDIX D

UNIFORM PRICE AUCTION (SHORTFALL STRATEGY)

As was shown in the text, the optimal bid price in the uniform price auction is always the reservation price of the bidder. In the case, of the shortfall strategy the bidder will set the bid price equal to the average costs of production at the bid quantity: $p_i\left(a_i^*\right) = \frac{C_i\left(a_i^*\right)}{a_i^*} = \theta_i$. The rationale for this bidding strategy is identical to that discussed in the text. There is no benefit to the bidder from deviating from his true reservation value.

The choice of abatement strategy will affect the quantity of abatement that is bid by the nonpoint source. The utility maximizing quantity of abatement under the surplus abatement strategy is determined by solving the following expected profit maximization problem:

$$\begin{aligned} &\underset{a_i,\lambda_i}{Max} \quad \omega\bar{p}_{m+1}\lambda_i - C_i\left(a_i\right) \\ &s.t. \qquad \lambda_i = a_i(1+e) \\ &\qquad\qquad \lambda^{\max} = a^{\max}(1+e) \\ &\qquad\qquad \lambda_i \le \lambda'^{\max} \end{aligned}$$

This maximization problem can be simplified as:

$$\underset{a_i}{Max} \quad \omega\bar{p}_{m+1}a_i(1+e) - C_i(a_i)$$

$$s.t. \qquad a_i = a^{\max}(1+e)$$

$$H = \omega\bar{p}_{m+1}a_i(1+e) - C_i(a_i) + k_1(a^{\max}(1+e) - a_i)$$

Show first order conditions with slack constraints:

$$\frac{\partial H}{\partial a_i} \equiv \omega\bar{p}_{m+1} - \frac{\partial C_i(a_i)}{\partial a_i}\frac{1}{(1+e)} - k_1 = 0$$

$$\begin{cases} k_1 > 0 \\ a_i = a^{\max} \end{cases} \qquad or \qquad \begin{cases} k_1 = 0 \\ a_i < a^{\max} \end{cases}$$

If the expected contract price is greater than the marginal cost of abatement, which is adjusted to account for the Nash abatement strategy, at the maximum allowable bid quantity ($a^{\max}$), the nonpoint maximizes expected profit by bidding this amount. Otherwise, the nonpoint source will bid the quantity(a_i^*) where the "stop-out" price equals the adjusted marginal cost of abatement. The optimal bid quantity is the expected profit maximizing level: $\lambda_i^* \min[a^{\max}, a_i^*]$.

The uniform price auction serves as a truth revelation mechanism, as the optimal bid price is equal to the average cost of abatement at the optimal bid quantity. The optimal bid quantity is less than the first-best profit maximizing level as a result of informational rents from asymmetric information. The nonpoint source accounts for its abatement strategy of over-production in its selection of its bid quantity. However, the contract does ensure that each team member will produce individual abatement levels equal to their bid quantities. The optimal bidding strategy is depicted graphically in Figure D.1.

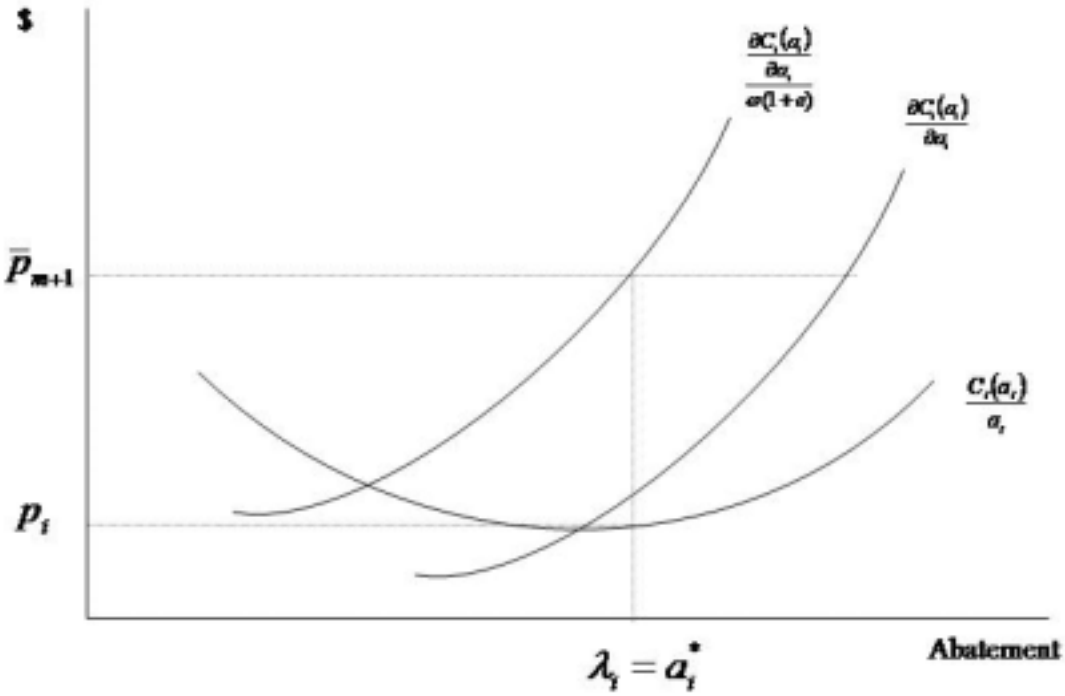

Figure D.1: Optimal Bid Price and Quantity for Uniform Price Auction and Shortfall Abatement Strategy

APPENDIX E

DISCRIMINATING PRICE AUCTION (SHORTFALL STRATEGY)

In a discriminating price auction, the nonpoint source following a shortfall Nash abatement strategy will bid $p_i = \dfrac{\dfrac{C_i(a_i^*)}{a_i^*}\dfrac{1}{(1+e)} + \dfrac{(1-F_{(m)}^{-i})}{f_{(m)}^{-i}}}{\omega}$ and $\lambda_i = \min\left[a^{\max}, a_i^*\right]$.

The nonpoint source maximizes expected profit, where the expectation is now based on the uncertainty associated with being selected into the team[22], represented by the distribution of the mth order statistic:

$$\underset{p_i, a_i, \lambda_i}{Max}\left(1-F_{(m)}^{-i}\right)\left(\omega p_i \lambda_i - C_i(a_i)\right) + F_{(m)}^{-i}(0)$$

$$s.t. \qquad a_i = \frac{\lambda_i}{(1+e)}$$

$$\lambda_i \le \lambda^{\max}$$

$$a^{\max}(1+e) = \lambda^{\max}$$

The maximization problem can be simplified as:

$$\underset{p_i, a_i, \lambda_i}{Max}\left(1-F_{(m)}^{-i}\right)\left(\omega p_i \lambda_i - C_i(a_i)\right) + F_{(m)}^{-i}(0)$$

$$s.t. \qquad a_i(1+e) = \lambda_i$$

$$\lambda_i \le a^{\max}(1+e)$$

[22] Weather uncertainty is incorporated into the choice of abatement strategy.

Finally, the maximization problem can be written in terms of actual abatement, a_i:

$$\underset{p_i, a_i}{Max}\left(1-F_{(m)}^{-i}\right)\left(\omega p_i a_i(1+e)-C_i\left(a_i\right)\right)$$

$$s.t. \qquad a_i \le a^{\max}$$

Solving the nonpoint source profit maximization problem to determine optimal bidding strategy will give:

$$H=(1-F_{(m)}^{-i})\left(\omega p_i a_i(1+e)-C_i\left(a_i\right)\right)+k_1(a^{\max}-a_i)\,.$$

This gives the following first order conditions:

$$\frac{\partial H}{\partial p_i} \equiv -f_{(m)}^{-i}\left[\omega p_i a_i(1+e)-C_i\left(a_i\right)\right]+(1-F_{(m)}^{-i})a_i\omega(1+e)=0$$

$$\frac{\partial H}{\partial a_i} \equiv (1-F_{(m)}^{-i})\left[\omega p_i(1+e)-\frac{\partial C_i\left(a_i\right)}{\partial a_i}\right]-k_1=0$$

$$\begin{cases} k_1>0 \\ a_i=a^{,\max} \end{cases} \qquad or \qquad \begin{cases} k_1=0 \\ a_i<a^{\max} \end{cases}$$

Rewriting the first order condition taken with respect to p_i condition shows the optimal price bidding strategy for the surplus abatement strategy under the discriminating price auction: $p_i = \frac{C_i\left(a_i^*\right)}{a_i^*}\frac{1}{\omega(1+e)}+\frac{(1-F_{(m)}^{-i})}{f_{(m)}^{-i}}$. Each bidder will inflate their bid above the reservation price extracting informational rent. The term $\frac{(1-F_{(m)}^{-i})}{f_{(m)}^{-i}}$ is the typical hazard function commonly found in problems of adverse selection. The numerator is the probability of being selected into the team conditional on the bidder's reservation price, and the numerator is the change in the probability of being selected into the team

corresponding to a unit increase in bid price above the reservation price. The greater the impact an increase in bid price has on the probability of team entry, the smaller the overall informational rent. Therefore, the nonpoint sources at the low-cost end of the distribution will extract greater informational rents than nonpoint sources at the high-cost end of the distribution. The optimal bid price line is always greater than the average abatement cost ensuring that the individual rationality constraint is always met given the Nash abatement strategy.

The remaining first order condition provides the optimal quantity bidding strategy. Rewriting the equation in terms of bid price, p_i, gives:

$$\frac{\partial H}{\partial a_i} \equiv (1 - F_{(m)}^{-i})\left[p_i - \left(\frac{\partial C_i(a_i)}{\partial a_i} \frac{1}{\omega(1+e)} \right) \right] - \frac{k_1}{\omega(1-e)} = 0 .$$

By substituting the optimal bid price line for p_i into the equation, the optimal bid quantity can be represented in terms of the relationship between the optimal bid price line and the marginal cost of actual abatement (i.e., the marginal cost of abatement adjusted to account for the optimal abatement strategy).

$$\frac{\partial H}{\partial a_i} \equiv (1 - F_{(m)}^{-i})\left[\left(\frac{C_i(a_i)}{a_i} \frac{1}{\omega(1+e)} + \frac{(1 - F_{(m)}^{-i})}{f_{(m)}^{-i}} \right) - \left(\frac{\partial C_i(a_i)}{\partial a_i} \frac{1}{\omega(1+e)} \right) \right] - \frac{k_1}{\omega(1+e)} = 0$$

$$\begin{cases} k_1 > 0 \\ a_i = a^{\max} \end{cases} \quad or \quad \begin{cases} k_1 = 0 \\ a_i < a^{\max} \end{cases}$$

When the optimal bid price line is greater than the marginal cost of abatement at the maximum bid quantity, the nonpoint source bids the maximum. Otherwise, the nonpoint source will bid the expected profit maximizing quantity, where the optimal bid price line intersects the marginal cost of abatement curve (Figure E.1).

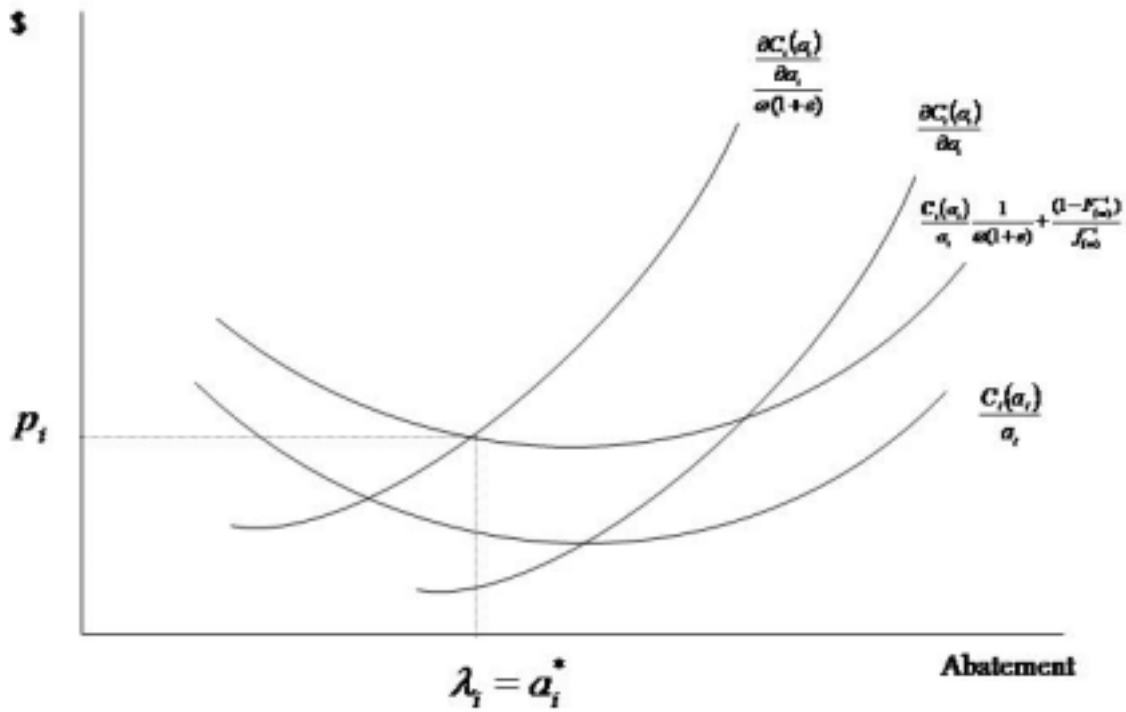

Figure E.1: Optimal Bid Price and Quantity for Discriminating Price Auction and Shortfall Abatement Strategy

BIBLIOGRAPHY

Alchian, Armen A. and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization" American Economic Review 62(5): 777-795.

Apogee Research, Inc. Incentive Analysis for Clean Water Act Reauthorization: Point Source/Nonpoint Source Trading For Nutrient Discharge Reductions. Report Prepared for Office of Water, and Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency. April, 1992.

Atkinson, Scott and Tom Tietenberg. 1991. "Market Failure in Incentive-Based Regulation: The Case of Emissions Trading" Journal of Environmental Economics and Management. 21, 17-31

Ausubel, Lawrence M. and Peter Cramton. 2002. "Demand Efficiency and Inefficiency in Multi-Unit Auctions" Department of Economics Working Paper, University of Maryland, College Park, MD.

Baron, John M. and Kathy Paulson Gjerde. 1997. "Peer Pressure in an Agency Relationship" Journal of Labor Economics. 15(2): 234-254.

Beavis, Brian and Martin Walker. 1983. "Achieving Environmental Standards with Stochastic Discharges" Journal of Environmental Economics and Management 10, 103-111.

Burtraw, Dallas. 1996. "The SO2 Emissions Trading Program: Cost Savings Without Allowance Trades" Contemporary Economic Policy. Vol. 14 pp. 79-94.

Bystrom, Olof and Daniel W. Bromley. 1998. "Contracting for Nonpoint-Source Pollution Abatement" Journal of Agricultural and Resource Economics 23(1):39-54.

Cabe, Richard and Joseph A. Herriges. 1992. "The Regulation of Non-Point Source Pollution Under Imperfect and Asymmetric Information" Journal of Environmental Economics and Management 22: 134-146

Coase, Ronald. 1960. "The Problem of Social Cost" Journal of Law and Economics. 3(1): 1-44

Crocker, Thomas D. 1966. "The Structuring of Atmospheric Pollution Control Systems" in, The Economics of Air Pollution. ed. Harold Wolozin.

Dales, J.H. 1968a. Pollution, Property, and Prices. University of Toronto Press. Toronto, Canada.

_____. 1968b. "Land, Water, Ownership" Canadian Journal of Economics. 1(November): 797-804.

Davies, J. Clarence and Jan Mazurek, 1997. Regulating Pollution. Does the U.S. System Work? Resources for the Future. Washington, DC

Eheart, Wayland J., Erhard E. Joeres, and Martin H. David. 1981. "Distribution Methods for Transferable Discharge Permits" Water Resources Research. Vol. 16. No. 2. (April) pp. 263-270.

Ellis, Ralph and Tracienne Ravita. 1997. "Scientific Uncertainties, Environmental Policy, and Political Theory" The Philosophical Forum. 28(3):209-231

Feinberg, Joel. 1965. "The Expressive Function of Punishment" The Monist. 49: 397-423.

_____. 1984. The Moral Limits of the Criminal Law Volume One: Harm To Other. Oxford University Press

Hahn, Robert W. 1989. "A New Approach to the Design of Regulation in the Presence of Multiple Objectives" Journal of Environmental Economics and Management. 17, 195-211

Harris, Milton and Artur Raviv. 1981. "Allocation Mechanisms and the Design of Auctions" Econometrica 49(6): 1477-1499.

Holmstrom, Bengt. 1982. "Moral hazard in teams" Bell Journal of Economics 13: 324-340

Horan, Richard D. 2001. "Difference in Social and Public Risk Perceptions and Conflicting Impacts on Point/Nonpoint Trading Ratios" American Journal of Agricultural Economics 83(4), 934-941.

Huanxin, W., Presley, B.J., and Velinsky, D.J. 1997. "Distribution and Sources of Phosphorus in Tidal River Sediments in the Washington, DC, Area" Environmental Geology. 30(3): 224

Kagan, Shelly. 1998. Normative Ethics. Westview Press. Boulder, CO

Kandel, Eugene, and Edward P. Lazaer. 1992. "Peer Pressure and Partnerships" Journal of Political Economy. 100(4): 801-817.

Karp, Larry. 2002. "Nonpoint Source Pollution Taxes and Excessive Tax Burden" Department of Agricultural and Resource Economics Working Paper, University of California, Berkeley, Berkeley, CA.

Krupnick, Alan J., Wallace E. Oates, and Eric Van De Verg. 1983. "On Marketable Air-Pollution Permits: The Case for a System of Pollution Offsets" Journal of Environmental Economics and Management. 10, 233-247

Letson, David. 1992. "Point/Nonpoint Source Trading of Pollution: An Interpretative Survey." Natural Resources Journal. 32: 233-247.

Lichtenberg, Eric and David Zilberman. 1988. "Efficient Regulation of Environmental Health Risks" Quarterly Journal of Economics Vol. 103, Issue 1, 167-178.

Licthenberg, Eric, David Zilberman, and Kenneth T. Bogen. 1989. "Regulating Environmental Health Risks under Uncertainty: Groundwater Contamination in California" Journal of Environmental Economics and Management 17, 22-34.

Malik, Arun S. 1990. "Markets for Pollution Control when Firms are Noncompliant" Journal of Environmental Economics and Management. 18, 97-106.

Malik, Arun A., David Letson, and Stephen R. Crutchfield. 1993 "Point/Nonpoint Source Trading of Pollution Abatement: Choosing the Right Trading Ratio" American Journal of Agricultural Economics 75, 959-967.

Malik, Arun S., Bruce A. Larson, and Marc Ribaudo. 1994. "Economic Incentives for Agricultural Nonpoint Source Pollution Control" Water Resources Bulletin. Vol. 30, No. 3 (June) pp. 471-480.

McAfee, R. Preston and John McMillan. 1991. "Optimal Contracts for Teams" International Economic Review 32(3): 561-577.

McGartland, Albert. 1988. "A Comparison of Two Marketable Discharge Permit Systems" Journal of Environmental Economics and Management. 15, 35-44.

McGartland, Albert M., and Wallace E. Oates. 1985. "Marketable Permits for the Prevention of Environmental Deterioration" Journal of Environmental Economics and Management. 12, 207-228.

Mendelsohn, Robert. 1986. "Regulating Heterogeneous Emissions" Journal of Environmental Economics and Management. 13 (December): 301-312.

Meran, Georg and Ulrich Schwalbe. 1987. "Pollution Control and Collective Penalties" Journal of Institutional and Theoretical Economics 143(4): 616-629.

Montgomery, David W. 1972. "Markets in Licenses and Efficient Pollution Control Programs" Journal of Economic Theory. 5, 395-418.

Noll, Roger, G. 1982. "Implementing Marketable Emissions Permits" American Economic Review. Vol. 72. No. 2. pp.120-124. May.

Organization for Economic Co-operation and Development. 1992. The Polluter Pays Principle: OECD Analyses and Recommendations (OECD/GD(92)81). Paris.

Page, Talbot. 1986. "Responsibility, Liability, and Incentive Compatability" Ethics. 97(1):240-262

Philips, Michael. 1985. "The Inevitability of Punishing the Innocent" Philosophical Studies. 48: 389-391.

Pushkarskaya, Helen. 2003. Nonpoint Source Water Pollution Control: Incentives Theory Approach. Unpublished PhD Dissertation. Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, OH

Griffin Ronald, and Daniel Bromley. 1982. "Agricultural Runoff as a Nonpoint Externality" American Journal of Agricultural Economics. 64: 547-552

Rasmussen, Eric. 1987. "Moral hazard in risk-averse teams" RAND Journal of Economics 18(3): 428-435.

Ruhl, J.B. 2000. "Farms, Their Environmental Harms, and Environmental Law" Ecology Law Quarterly. 27: 263-350

Sandmo, Agnar. 1975. "Optimal Taxation in the Presence of Externalities" Scandinavian Journal of Economics, 77: 86-98

Scher, George. 1984. "Right Violations and Injustices: Can We Always Avoid Tradeoffs?" Ethics 94(2): 212-224

Segerson, Kathleen. 1988. "Uncertainty and Incentives for Nonpoint Pollution Control" Journal of Environmental Economics and Management 15: 87-98

_____. 1990. "Liability For Groundwater Contamination from Pesticides" Journal of Environmental Economics and Management. 19(3): 227-243

Shortle, James S. 1990. "The Allocative Efficiency Implications of Water Pollution Abatement Cost Comparisons" Water Resources Research Vol. 25, Issue 5, 793-797.

Sohngen, Brent and Michael A. Taylor. 1998. "Incentive-Based Conservation Policy and the Changing Role of Government" American Farmland Trust Center for the Environment Working Paper Series.

Taff, S.J., and Senjem, N. 1996. "Increasing Regulators' Confidence in Point-nonpoint Pollutant Trading Schemes". Water Resources Bulletin. 32(6): 1187.

Tietenberg, T.H. 1985. Emissions Trading: an exercise in reforming pollution policy Resources for the Future, Inc. Washington, D.C.

United Nations Environment Program. 1992. Rio Declaration on Environment and Development. <http://www.unep.org/unep/rio.htm> (7 May 1998)

United States Code. Clean Water Act 33 U.S.C. 1251 et seq.

United States Environmental Policy Agency. 2002. Water Quality Inventory 2000 Report. Office of Water. Washington, DC (EPA-841-R-02-001)

_____ . 2003. Water Quality Trading Policy. Office of Water, Washington, DC

Wilson, Robert. 1979. "Auction of Shares" Quarterly Journal of Economics 94: 675-689

Woodward, Richard T. and Ronald A. Kaiser. 2002. "Market Structures for U.S. Water Quality Trading" Review of Agricultural Economics. 42(2): 366-383.

World Resources Institute. 2003. "Nutrient Net" < http://www.nutrientnet.org />

Xepapadeas, A.P. 1991. "Environmental Policy under Imperfect Information: Incentives and Moral Hazard" Journal of Environmental Economics and Management 20: 113-126