The Pennsylvania State University

The Graduate School

Department of Agricultural Economics and Rural Sociology

THREE ESSAYS ON PRICING IN SOCIALLY-OPTIMAL MARKETS FOR DIFFERENTIATED GOODS

A Dissertation in

Agricultural, Environmental and Regional Economics

by

Gaurav Somenath Ghosh

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2009

The dissertation of Gaurav Somenath Ghosh was reviewed and approved[1] by the following:

James S. Shortle
Distinguished Professor of Agricultural and Environmental Economics
Dissertation Co-Advisor
Chair of Committee

Anthony M. Kwasnica
Associate Professor of Business Economics
Dissertation Co-Advisor

Edward C. Jaenicke
Associate Professor of Agricultural Economics
Graduate Program Chair

Richard Ready
Associate Professor of Agricultural and Environmental Economics

Stephen M. Smith
Professor of Agricultural and Regional Economics
Head of the Department of Agricultural Economics and
Rural Sociology

[1]Signatures on file in the Graduate School.

Abstract

In my first essay I report results from an analysis where representative methods of estimation from the Classical and Bayesian approaches to statistical inference are empirically compared. The chosen Classical methods are based on Least Squares and Maximum Likelihood and the chosen Bayesian method is the Hierarchical model. Each method is applied to a spatial hedonic property value model. The resulting estimates are then compared using nonparametric tests. The comparisons are then used to make inferences on the relative accuracy, precision and quality of the different methods of estimation. The Hierarchical Bayesian and Classical Maximum Likelihood methods are found to supply the best estimators of the spatial regression model and predictors of house price. Significant differences are also found in the relative accuracy and precision of the methods. I infer that the Hierarchical Bayesian and Classical Maximum Likelihood methods are best suited to prediction and estimation of spatial hedonic property value models.

In my second essay I compare two tradable permit markets in their ability to meet a stated environmental target at least cost when some polluters have stochastic and non-measurable emissions. The environmental target is of the safety-first type, which requires probabilistic control of emissions. One market is built around the trading ratio, which defines the substitution rate between stochastic and deterministic pollution, and is modeled on existing markets for water quality trading. The other market is built around a new definition of the traded commodity as a multi-attribute good, where the attributes supply information to the market on the environmental risks associated with stochastic pollution. The latter market is found to out-perform the trading ratio market in its ability to satisfy the safety-first environmental target at least cost. This result comes about because polluters are able to directly price risk in the latter market. In the former market risk is not a factor in the trading decision and can only be controlled under highly restrictive conditions.

In my third essay I report results from an economic experiment where the two markets developed in the previous essay are compared in a testbed that captures important features of existing markets for water quality trading. In the interests of tractability these features were abstracted from in the previous theoretical analysis. One feature is that of oligopsony and the second feature is that of a discrete trading environment where polluters generate credits by implementing one of a small set of emission-reducing technologies. The experimental results indicate that the market with multi-attribute goods generates a superior environmental outcome to the trading ratio market. Furthermore, the average cost of pollution control is lower in the former market. Market power is independent of the type of market institution, but I do find that large buyers have more market power than small buyers. Finally, I find that sellers of credits learn to resist market power as they gain experience, but at the cost of market efficiency since their resistance causes a fall in the number of trades. Overall, the results support the thesis of the second essay, that the market with multi-attribute goods generates better environmental outcomes than the trading ratio market.

Table of Contents

List of Tables

List of Figures

Acknowledgments

Conducting my research and writing this dissertation has been an osmotic process. I have learned much about the topics in my dissertation, the research process and life in general from members of my committee, faculty at Penn State and elsewhere, the many friends I made in State College, and last but not least, my ever supportive wife and family. Now at the end of this long, interesting and grueling process, acknowledgements are in order.

Thanks to my committee for their time and advice, and for gamely navigating my often incomprehensible thickets of prose to find, more often than not, the few kernels of sense and value. To Rich Ready for his trenchant commentary on the weaknesses in my method and analysis. To Ted Jaenicke for willingly sharing his insights into the academic profession. To Tony Kwasnica for helping me develop, execute and understand the economic experiment that formed the basis of my third essay. And finally to Jim Shortle for guiding me with an unerring hand, for giving me the freedom to develop my research in sometimes unexpected directions, and for his unwavering support in all aspects of my career at Penn State. I would also like to thank Murali Haran for sparking my interest in spatial statistics and overseeing the earliest iteration of my first essay, Ann Fisher for her example and guidance during my first year, Alessandro Bonanno for his perspective on newly minted academic careers, and everyone who let me attend their lectures.

Thanks to my friends in the AEREC program. To Fernando for sharing his data set and many beery hours. To Nga for being the only constant presence in the transient world that is graduate school. To Greg for being my econometric encyclopedia. To Kelly who showed through example that research is just another nine-to-five job. To Amal for always being game for anything non-academic. To Debdatta for letting me inflict my Bengali on her. To Georg for always being good for a chat. To Hwansoo for getting me interested in Korean food. To Simanti for always being in good spirits. To Xiaohua for his infectious enthusiasm in the academic life. To David for letting me beat him in squash. To Hernan for providing references on demand. To Rodrigo for being ever helpful without complaint. To Milagro for always making sense. And thanks to everyone else I got to know at Penn State. You will not find your name here, but I will cherish the memories.

I also take this opportunity to thank my family, without whom I would never have started down the path that led to these essays. To Sharmila Ghosh for teaching me to not take life too seriously and to always live it to the fullest. To Somenath Ghosh for always knowing the correct path, but also giving me the freedom to make my mistakes and learn. To Nirmal and Chandra Bose for giving me a second home here in State College. To Aroop and Sucharita Ghosh for feeding me and letting me vegetate when I visit. To Anna, Meena and Bonnie for dispensing advice as only older cousins can. To Abhik the optimist who pretends to expect the worst. To Siddhartha for showing me how life should be lived. And finally, but most importantly, to Sushma for loving me, marrying me and believing in me even when I doubted.

Try to learn something about everything and everything about something.

–Thomas H. Huxley

Reading made Don Quixote a gentleman. Believing what he read made him mad.

–George Bernard Shaw

Rome did not create a great empire by having meetings, they did it by killing all who opposed them.

–Unknown

Chapter 1

Bayesian and Classical Inference in the Context of Spatial Hedonic Modeling

"Location, location, location" is perhaps the most hackneyed phrase in a realtor's vocabulary, but its truth cannot be denied. Property values are affected by the attributes of other properties in the vicinity, and by neighborhood amenities and disamenities . People value a house near an airport and in the flight path less than a similar house further away (Feitelson et al. 1996). A house a stone's throw from a landfill costs significantly less than an identical house at a further location (Nelson et al. 1992). Since geography matters it is reasonable to expect spatial hedonic econometric models of property markets to be theoretically and empirically superior to hedonic models that abstract from spatial processes (Brasington and Hite 2005; Cohen and Coughlin 2007). When space is neglected statistical estimates become biased and inefficient because of the effects of omitted variables and model misspecification (Anselin 1988). A single observation carries less information when property prices are positively correlated than when prices are independent because price at a location is partially explained by prices at neighboring locations. Positive correlation reduces the effective size of the sample (Schabenberger and Gotway 2005, pg. 34). Speciously assuming spatial independence results in distortion of the sampling variance of the parameter estimates, which modifies the shape of the sampling distribution and thus the probabilities of Type I and Type II errors (Griffith

1988, pg. 82). Hypothesis testing, statistical significance and confidence intervals are affected.

There is a broad literature on methods for unbiased and efficient estimation and prediction under conditions of spatial correlation. In the hedonic literature spatial correlation is modeled either through incorporation into the trend component or into the error component of the model.[1] It is incorporated into the trend (error) component by making the dependent variable a function of a spatially weighted average of the known (unknown) characteristics of the neighboring properties (Anselin 1988). In the error component case spatial correlation is assumed to be a function of omitted variables (Gelfand et al. 2004; Kim et al. 2003). Trend-based spatial models are commonly referred to as spatial lag models and error-based spatial models as spatial error models. In both cases spatial correlation is modeled through a weights matrix, which defines the correlation between all pairs of location in the data set. Following Tobler's First Law of Geography – "everything is related to everything else, but near things are more related than distant things" – correlation is generally modeled as an inverse function of distance (Tobler 1970).

Methods for estimation and prediction under spatial correlation can be placed into two distinct categories. The first set of methods is based on the Classical approach to statistical inference, and the second set on the Bayesian approach. The Classical approach is based on a non-probabilistic methodology as formulated in seminal texts like

[1] Consider a standard econometric model of the form $y = X\beta + \epsilon$ where y is the dependent variable, X the vector of explanatory variables, β the vector of associated coefficients and ϵ the error term. I refer to $X\beta$ as the trend component and ϵ as the error component.

Fisher (1925, 1935) and Neyman and Pearson (1933). It finds its philosophical underpinning in the premise that inference is objective, universal and not conditional on the data (Poirier 1988). The Bayesian approach uses an explicitly probabilistic methodology and operates under the premise that inference is necessarily data-conditional and subject to the researcher's belief structure (Howson and Urbach 1991). The Classical approach was borne out of criticisms of Bayesian subjectivity and its place in science in the 1920s and 1930s, but in subsequent decades these criticisms have been largely refuted (e.g. Edwards et al. 1963). For exploration of the foundational and philosophical differences between the Classical and Bayesian approaches, the reader may find Bernardo and Smith (1994) and Jeffreys (1961) great places to start.

Today, even though Classical methods predominate, Bayesian estimation methods are extensively applied in fields as disparate as geography, bio-statistics and ecology. Methodological choice has become largely a matter of familiarity and problem context, and Bayesian and Classical methods are often seen as close substitutes. Yet, Cressie (1993) warns that the foundational differences between the two approaches impact the interpretation and quality of estimates, predictions and inferences, as measured by criteria like unbiasedness, efficiency and robustness. In the context of property market analysis, methodological choice may affect implicit price estimates of housing attributes, property valuations and feasibility studies for new housing and infrastructure developments.

Some recent Bayesian hedonic property market analyses (e.g. Clapp et al. 2002; Gelfand et al. 2004) mark the entry of Bayesian methods into a field dominated by Classical methods. Given this trend a comparison of Bayesian and Classical methods

specifically focussed on the quality of their results when applied to property markets will be timely and instructive. At the very least such a study may help property market researchers with methodological choice given the context of their problem. This paper provides such a comparison. Knowing the quality of a method is important because it affects the reliability of the attendant inferences and prescriptions. Uncertainty in results implies uncertain inference and uncertainty in the consequences of policy prescriptions, which is undesirable from a policy analysis perspective (Beron et al. 2004).

I use a large sample approach to compare the performance of representative estimation methods from the Bayesian and Classical schools when applied to a spatial hedonic model of a property market. The Classical methods chosen are Least Squares and Maximum Likelihood and the Bayesian method is Hierarchical Bayesian. Each method is applied to the spatial hedonic model to generate estimates of the model parameters. The estimates are compared through appropriate metrics to gauge their quality. Quality estimates will be accurate and precise. Accuracy is determined through a measurement of bias: accurate estimators have low bias. Precision is measured by studying the dispersion of the estimates and is inversely related to variability. By comparing the quality of the estimates, inferences are drawn on the quality of the corresponding estimators and methods. Since price prediction at other locations in the property market area is often important, the predictive capabilities of the methods are tested through analogous metrics.

The results indicate that a nuanced approach to the choice of estimation method is best. Different methods performed best in different contexts. Some methods were more accurate while others were more precise. Some methods were better at estimating model

parameters and others were better at prediction. For example, I find that the Hierarchical method was unequivocally superior at estimating the structure of spatial correlation, but the Maximum Likelihood method was best at estimating the trend coefficients. The Least Squares estimators, in contrast, performed relatively poorly. The results are discussed in detail in the final two sections of the paper.

The paper is structured as follows. First, using hedonic price theory an economic model is built of the property sales market in Bogotá, Colombia. Second, the data set is described and descriptive statistics provided. Third, a sampling strategy that enables inference through the use of large sample statistical theory is discussed. Fourth, a spatial error econometric model is developed and the Least Squares, Maximum Likelihood and Hierarchical Bayesian estimators described. Fifth, estimates and predictions from the different methods are collected, compared and ranked. The paper concludes with a discussion of the salient results.

Hedonic property value models are widely used in the analysis of property markets. These models utilize hedonic price theory, which formalizes the notion that a differentiated commodity is in essence a package of n attributes, $\boldsymbol{x} = (x_1, \cdots, x_n)$ and its price $p(\boldsymbol{x})$ is a function of the levels of these attributes. When sufficiently large numbers of the differentiated commodity are available such that consumers can choose from a wide variety of alternative packages, hedonic price theory shows that at the market equilibrium $p(\boldsymbol{x})$ equals each consumer's bid or value function for the differentiated commodity $\xi(\boldsymbol{x})$ (Rosen 1974). Implicit market valuation of attributes becomes possible since individual valuations are equalized at the margin and set to the marginal price at the market equilibrium.

Consider a representative utility-maximizing consumer with income y. Her utility function is $U(\boldsymbol{x}, z)$ where $\boldsymbol{x}$ is the differentiated commodity and z is a composite commodity representing all other commodities consumed. When price of z is normalized to one, assuming that the consumer buys one unit of $\boldsymbol{x}$ her budget constraint is $y = z + p(\boldsymbol{x})$. The consumer maximizes her utility by choosing z and $\boldsymbol{x}$ such that first order conditions (1.1) and her budget constraint are met. The first order conditions are necessary and sufficient when U is convex and $p(\boldsymbol{x})$ is not too concave (Intriligator 1971; Rosen 1974).

$$\frac{\partial U/\partial x_i}{\partial U/\partial z} = \frac{U_{x_i}}{U_z} = p_i = \frac{\partial p}{\partial x_i} \quad \forall\, i = 1, \cdots, n \tag{1.1}$$

The consumer's bid function $\xi(\boldsymbol{x}; u, y)$ is her willingness to pay for $\boldsymbol{x}$ given her income and desired level of utility u. By definition $\xi(\boldsymbol{x}; u, y)$ satisfies $U(\boldsymbol{x}, y - \xi) = u$. Since $\xi_{x_i} = U_{x_i}/U_z$ it follows from (1.1) that $\xi_{x_i} = p_i$ for all i. The utility maximizing commodity $\boldsymbol{x}^*$ is chosen when bid and price surfaces are tangential and $p(\boldsymbol{x}^*) = \xi(\boldsymbol{x}^*)$. Rosen (1974) shows that simultaneous utility maximization for all consumers only occurs at the market equilibrium where every consumer's valuation of attribute x_i is p_i, commonly interpreted as the marginal implicit price of the attribute. The attributes consist of market and non-market goods.

Hedonic price theory is applied to property market analysis because properties are easily modeled as differentiated goods. Attributes consist of structural and neighborhood characteristics such as house size, the number of bedrooms, crime rates and the school system. The true impact of the attributes on house price is not known, but is typically approximated through a reduced form relationship between the sales price of a property

and its attributes. The sales price is assumed to equal the market equilibrium price. The coefficients in the reduced form are estimated and then interpreted as (some function of) the implicit price of the associated non-market good. The valuation of non-market goods when these goods are attributes of the properties traded in the market becomes feasible through the use of hedonic property value models.

For the analysis in this paper the relationship between property price and attributes is captured through a semi-log reduced form equation. The semi-log specification is commonly used in hedonic analysis (e.g., Lake et al. 2000; Baranzini and Ramirez 2005; Pope 2005) because it permits intuitive interpretation of the coefficients associated with the attributes, which serve as the independent variables or covariates in the model. The coefficients are interpreted as implicit price elasticities (semi-elasticities) when the covariates enter (do not enter) the regression equation in the log scale (Greene 2003, pg. 123) and as median impacts when the covariates are dummy variables (Gujarati 2003, pg. 320). Also, by using log prices, the residuals (and by inference the errors) are normalized and the effects of outliers are reduced. If the general form of the regression equation is $y = \boldsymbol{X}\beta + \epsilon$ then under the semi-log specification $y = \ln p(\boldsymbol{x})$ and $\boldsymbol{X} = \{\boldsymbol{x}', \ln \boldsymbol{x}''\}$ where $\boldsymbol{x}'$ ($\boldsymbol{x}''$) is the vector of attributes that do not enter (enter) the regression equation in the the log scale.

Early use of the hedonic method to study property markets include Tiebout (1956), who used it to study the implicit market for neighborhoods. The hedonic approach was first applied to environmental valuation by Ridker and Henning (1967) who found a strong statistical relationship between housing values and air quality in the

metropolitan area of St. Louis. Since then hedonic theory has been used to value environmental factors such as air quality (e.g., Kim et al. 2003; Smith and Deyak 1975; Zabel and Kiel 2000), water quality (e.g., Leggett and Bockstael 2000; Poor et al. 2001), agricultural externalities (e.g., Palmquist et al. 1997; Ready and Abdalla 2005), noise (e.g., Cohen and Coughlin 2007) and proximity to hazardous facilities (e.g., Brasington and Hite 2005; Nelson et al. 1992), through their impacts on property values. For further reading on hedonic price theory in the environmental context reference to Freeman (2003) and Taylor (2003) is recommended.

1.1 The Data

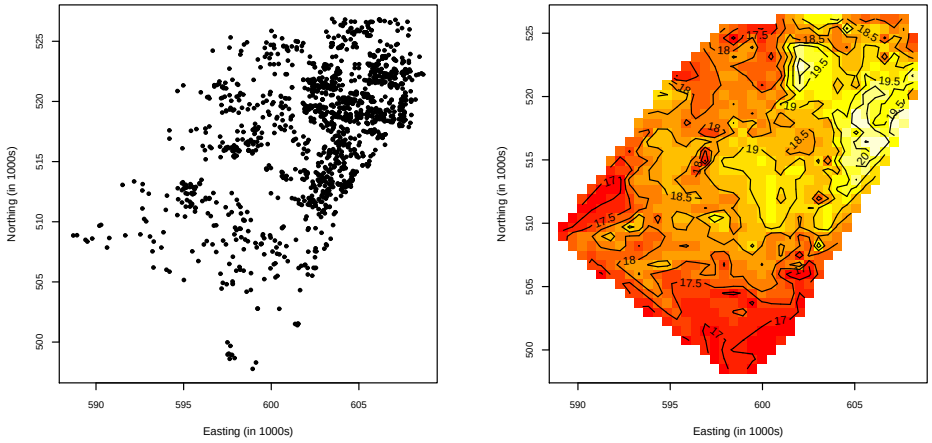
The data set was compiled by Fernando Carriazo Osorio as part of his dissertation research (see Carriazo Osorio 2007). I use the data with his kind permission. The data set consists of sales information on 1853 non-apartment housing units in Bogotá, Colombia. Each housing unit is a detached or semi-detached residential property. Data on sales price, location and structural attributes were provided by `Metrocuadrado.com`, a realtor based in Bogotá. All sales occurred between 2001 and 2006 and prices were deflated by the Colombian Consumer Price Index, with 2005 as the base year. The price index data were obtained from Colombian Central Bank official statistics.[2] Each housing unit is geo-referenced, originally in latitude and longitude, but converted into Bogotá Zone Transverse Mercator coordinates. Distortion due to the Earth's curvature is lower under the transverse Mercator projection than under standard representations of latitude and

[2] Available online at `http://www.banrep.gov.co/statistics/sta_prices.htm`

longitude. It allows for more accurate interpretation of spatial relationships denominated in Euclidean distance. Latitude is measured in Northings and longitude in Eastings.

The housing units cluster towards the northeast of the city, as seen in Figure 1.1(a), which as indicated by Figure 1.1(b) also contains the more expensive housing. The sample only contains records of sales in the formal housing market, which is concentrated in northeast Bogotá. The spatial distribution and price structure is explained by Bogotá's geography and history. The terrain in the northeast is hilly with a cooler

Fig. 1.1. Location and Prices of Properties



(a) Spatial Distribution (b) Contour Plot of *LnSale*

climate and attractive vistas. This made it attractive as a residential location drawing in upper income households over the years. These households conducted their property transactions under the purview of the formal housing market and generated official records of their deals, which could later be assembled into property market data. In contrast, the south and the west became industrial regions and were settled by lower income households. Much of the housing in these areas sprang up outside the formal

housing market and without formal records, data on housing transactions in these areas is patchy and inconsistent. The larger realtors like Metrocuadrado do not participate in these markets.

Descriptive statistics of the dependent variable, log of property price ($\ln p$), and the covariates are given in Table 1.1. *Northing* and *Easting* locate each property on the

Table 1.1. Variable List and Descriptive Statistics

Variable	Units	Mean	Std.Dev.	Min	Max
$\ln p$	$\ln(pesos)$	18.95	0.77	15.88	20.77
$Northing$	km	20.19	5.12	0.00	29.09
$Easting$	km	14.21	3.46	0.00	19.97
$\#Baths$	N/A	2.84	0.83	1	4
$\#Garages$	N/A	1.08	0.83	0	2
$\ln(Area)$	$\ln(m^2)$	5.27	0.60	3.33	6.91
$Carpet$	N/A	0.61	0.49	0	1
$DiningRoom$	N/A	0.59	0.49	0	1
$24DoorKeep$	N/A	0.38	0.49	0	1
$\ln(PM_{10})$	$\ln(\mu g/m^3)$	4.05	0.24	3.47	4.62
$\ln(ZonalPark)$	$\ln(m)$	7.04	0.80	3.91	8.42
$\ln(FloodArea)$	$\ln(m)$	7.38	1.10	3.91	8.87
$\ln(72^{nd}St)$	$\ln(m)$	8.57	1.07	3.91	9.84
$\ln(HistCent)$	$\ln(m)$	9.02	0.95	3.91	9.78
$CrimeIndex$	homicides / 100000	3.17	5.53	0	151
$Stratum$	N/A	4.18	1.14	1	6

map. The origin has been shifted such that the most southwest property in the sample has the coordinates $(0, 0)$. The sample is collected from a $30km \times 20km$ rectangle as indicated by the minimum and maximum values of *Northing* and *Easting*. Comparison of Figures 1.1(a) and 1.1(b) indicates correlation between property price and location and I expect $\ln p$ to increase in both location variables.

The next seven variables are site-specific and / or structural. $\#Baths$, $\#Garages$ and $\ln(Area)$ are self-explanatory. $Carpet$, $DiningRoom$ and $24DoorKeep$ are dummy variables, set to 1 when the house has carpeting, a dining room and 24 hour door-keeping services respectively, and 0 otherwise. Data on these variables were obtained from Metrocuadrado. The PM_{10} level measures the airborne concentration of suspended particulates of diameter of $10\mu g$ or less making $\ln PM_{10}$ a measure of air pollution. Locations with higher PM_{10} levels are more polluted. Pollution data were obtained from the Departmento Tècnico Administrativo del Medio Ambiente, the environmental authority in Bogotá. There are 11 monitoring stations in Bogotá, each geo-referenced. A raster of PM_{10} level was produced through inverse distance interpolation of the emissions recorded at each station. By superimposing the raster on the property map, PM_{10} concentrations at each location were estimated. I expect that $\ln p$ to increase with all site-specific variables except $\ln(PM_{10})$.

The next four variables measure the distance between the properties and (dis)amenities. Historical Center and 72^{nd} Street are the two central business areas in Bogotá. Zonal parks are small green spaces maintained by local planning committees. Under the assumption that households prefer to live close to the city center and to green spaces I expect the corresponding coefficients to be negative. Low-lying areas are designated as flood prone. The Flood Area variable is a proxy for elevation relative to the surrounding topography, but not absolute elevation. Properties in these areas are expected to command lower sales prices because of flood risk and because of poorer vistas implying a positive coefficient for $\ln(FloodArea)$. Data on the location of (dis)amenities were

provided by Bogotá's urban planning department. The amenities are geo-referenced and merged with properties into a single map.

The final two variables are defined at the neighborhood level. $Stratum$ is a count variable ranging from one to six and indexing average income in the neighborhood. The subset of neighborhoods with highest mean incomes have $Stratum = 6$. It is expected to trend positively with $\ln p$ and was obtained from Metrocuadrado. The final variable $CrimeIndex$ was obtained from the Unified System of Information on Violence and Delinquency and measures neighborhood crime rates. The crime data was geocoded by matching neighborhoods to individual properties. Under the hypothesis that households prefer to locate to low-crime neighborhoods property prices are expected negatively correlate to $CrimeIndex$.

1.2 A Case for Large Sample Based Inference

In a spatial error model, the errors at different locations are modeled as realizations of a random field, which is a stochastic process in two or more dimensions.

1.2.1 The Stationarity and Random Field Assumptions

The random field assumption implies that a sample of n locations is a *single* realization from an n-dimensional distribution (Banerjee et al. 2004, pg. 31). The researcher does not know if this single realization is a representative draw from the underlying population or an outlier. Inference about the underlying population is impossible except under the assumption that the random field is stationary (Schabenberger and Gotway 2005). Although almost universally used, the stationarity assumption is often untenable.

Also, even when stationarity does exist reliable inference is only feasible for trend-related parameters – the βs in the standard regression equation – and not for the error-related parameters that define the correlation structure of the random field. The distributional properties of error-related parameter estimates are not well understood under the random field assumption, which implies that their significance cannot be determined (Cressie 1993, pg. 99). Asymptotic variances of the error-related parameters, as calculated in Upton and Fingleton (1985) are of no use (Griffith 1988, pg. 25). These variances are for the limiting case $k \to \infty$ and should not be used if $k < 30$, where k is the number of realizations of the n-dimensional distribution.

1.2.2 Sampling Strategy

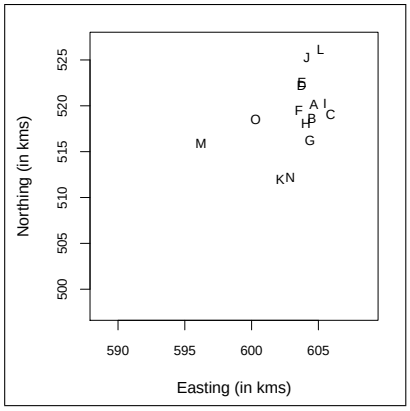
The suitability of the stationarity assumption and the lack of inference on error-related parameters under the random field assumption are problems that vanish when a large sample of n-dimensional realizations are drawn from the underlying population. As a consequence of the Law of Large Numbers using a large sample enables use of asymptotic sample theory for inference about model parameters.

Hence I propose to conduct a large sample analysis of the formal property market in Bogotá. The sample consists of $k = 38$ independent and identically distributed subsamples drawn randomly and without replacement from the original data set. Each subsample is of size $n = 100$, which implies that each subsample is a 100-dimension realization of the underlying stochastic process. Under the thumb rule that the sampling distribution of the sample mean of a non-normal parent distribution is approximately normal when the sample size is greater than 30 – in which case normality-based large

sample estimation theory is applicable – the 38 observation sample may be considered "large."

The Classical and Bayesian estimation methods are applied to each subsample and the parameter estimates recorded. The parameter estimates are then used to predict prices at 15 locations. The aim is to provide a flavor of the predictive capabilities of the methods of estimation and for that purpose a prediction data set of 15 observations is deemed suitable. The spatial distribution of these locations, indexed $\{A, B, \cdots, O\}$, is shown in Figure 1.2. The locations correspond to properties in the original data that were set aside for prediction and not included in the 38 subsamples. The spatial distribution of these set-aside properties approximates the distribution of the original data set. Prediction quality is gauged through direct comparison with the recorded sales

Fig. 1.2. Map of Prediction Points



prices of these properties. The 38 sets of estimations and predictions are analyzed to

enable inference about the relative quality of the estimation methods in the context of spatial hedonic modeling.

The proposed sampling scheme has similarities and differences with the bootstrap and the jackknife procedures, which though more comprehensive are infeasible in the current context. The bootstrap requires sampling with replacement, which means that an observation can manifest more than once in a subsample. As a result, the spatial configuration of the sample is distorted, which affects inference (Griffith 1988, pg. 28). The jackknife is not used because, given the size of the data set, it is too time consuming. The jackknife requires all subsamples of a given size to be analyzed. Given that $n = 1853$ if subsample size were 1852, then 1852 regressions are needed. If the subsample size were 1851, then 171,578 regressions are needed. Since Markov Chain Monte Carlo (MCMC) simulations using the entire sample – and needed for the Hierarchical method – take more than three days to run (see Ghosh and Carriazo-Osorio 2007), the jackknife is too time-consuming. The proposed sampling scheme may be likened to a *partial jackknife*. Like in the jackknife equal-sized subsets of the data are sampled, but unlike the jackknife, all available subsets are not sampled. The estimates are calculated in a manner analogous to that for the bootstrap or jackknife. They will have the same asymptotic properties as jackknife or bootstrap estimates, but will be less efficient due to the small number of sample points. The proposed approach is not as rigorous as the jackknife, but does allow recourse to large sample theory. Similar to the jackknife and bootstrap approaches, the econometric model remains unchanged across the k sets of regressions.

One should note that the problems with estimating error-related parameters under the random field assumption only affects the Classical estimation methods. The MCMC

algorithm used in the Hierarchical method easily simulates a sampling distribution of the autoregressive structure (Haining 1978). The sampling distribution tends towards the true distribution as the sample size increases implying that a good approximation of the standard error is obtainable even when asymptotic theory is unusable.

1.3 The Econometric Model

Since property values are affected by the attributes of neighboring properties, a spatial lag model is tenable. However the sample contains too few observations over the geographical window (i.e. Bogotá), which implies incomplete information about neighbors. The spatial lag model would result in a serious omitted variable problem, distorting results. This problem is avoided by using a spatial error model.

1.3.1 The Spatial Error Model

Let $\ln(Price)$ be a spatial process[3] $\{p(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{D} \subseteq \mathbb{R}^2_+\}$ where $\boldsymbol{s}$ is a generic location in $\boldsymbol{D}$, a fixed subset in two dimensional Euclidean space. $\boldsymbol{D}$ corresponds to the geographical window over which the stochastic process is defined (i.e. Bogotá). The spatial process $p(\boldsymbol{s})$ has a trend and error as shown in (1.2):

$$p(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta} + \epsilon(\boldsymbol{s}) \tag{1.2}$$

[3] As per convention a stochastic process in more than one dimension is called a spatial process (Banerjee et al. 2004, pg. 21).

where $\boldsymbol{X}(\boldsymbol{s}) \in \mathbb{R}^m$ is an m-vector of attributes or covariates, $\boldsymbol{\beta} \in \mathbb{R}^m$ is an m-vector of coefficients and $\epsilon(\boldsymbol{s})$ is the error. As noted in the first footnote $\boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta}$ is the trend component or mean structure of the regression equation. $\epsilon(\boldsymbol{s})$ is assumed to be distributed normally, which implies that a set of errors $\boldsymbol{\epsilon}(\boldsymbol{s}) = \{\epsilon(\boldsymbol{s}_1), \cdots, \epsilon(\boldsymbol{s}_n)\}$ corresponding to any set of locations $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n\}$ is distributed multivariate normal, $\boldsymbol{\epsilon}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ (and $p(\boldsymbol{s}) \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$). The normality assumption simplifies the distribution theory beyond what may be expected with other distributional assumptions (Banerjee et al. 2004, pg. 31). It is justified by the Central Limit Theorem, which states that the sum of many possibly non-normal processes is approximately normal. The error is decomposed into two components

$$\epsilon(\boldsymbol{s}) = w(\boldsymbol{s}) + \eta(\boldsymbol{s}) \tag{1.3}$$

where $w(\boldsymbol{s})$ is the measurable spatial effect and $\eta(\boldsymbol{s})$ is the nugget effect. The nugget effect captures measurement error and microscale variation, which is spatial correlation over distances smaller than $\min\{\| \boldsymbol{s}_i - \boldsymbol{s}_j \| : 1 \leq i, j \leq n\}$ and cannot be measured. $w(\boldsymbol{s})$ is modeled as a zero-centered, stationary normal process and $\eta(\boldsymbol{s})$ as an uncorrelated pure error term. For some set of locations $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n\}$, they are distributed multivariate normal with $\boldsymbol{w}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{H}(\boldsymbol{\phi}))$ and $\boldsymbol{\eta}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \tau^2\boldsymbol{I})$. Since $\boldsymbol{\epsilon}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, from (1.3)

$$\boldsymbol{\Sigma} = \sigma^2\boldsymbol{H}(\boldsymbol{\phi}) + \tau^2\boldsymbol{I} \quad \text{where } [\boldsymbol{H}(\boldsymbol{\phi})]_{ij} = \rho(\boldsymbol{\phi}; \boldsymbol{s}_i - \boldsymbol{s}_j) \tag{1.4}$$

where τ^2 is a parameter capturing variability because of nonspatial error, σ^2 is a parameter affecting the magnitude of the spatial correlation, $\rho(\boldsymbol{\phi}; \boldsymbol{h})$ is a covariance function

that defines the structure of correlation and how it decays over distance and $\boldsymbol{\phi}$ is a parameter vector. In the spatial statistics literature τ^2 and σ^2 are called the *nugget* and *partial sill* respectively.

The appropriate functional form of the spatial covariance function $\rho(\cdot)$ is determined through variogram analysis. The variogram measures the variance in the error as a function of distance or spatial lag and is $2\gamma(\boldsymbol{h};\boldsymbol{\theta}) = \text{var}\left(\epsilon(\boldsymbol{s}+\boldsymbol{h}) - \epsilon(\boldsymbol{s})\right)$, where $\boldsymbol{h}$ is the distance between two generic locations, $\boldsymbol{\theta}$ is the vector of trend and error coefficients and $\epsilon(\boldsymbol{s})$ is the error. If $\boldsymbol{\alpha} = \{\tau^2, \sigma^2, \boldsymbol{\phi}\}$ is the vector of error coefficients [see (1.4)] then $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$.

A number of variogram estimators have been proposed. The classical estimator is based on the variance of the squared difference between two points (Matheron 1962) but is sensitive to outliers. I use a robust alternative (1.5) based on the square root of the difference and proposed by Hawkins and Cressie (1984).

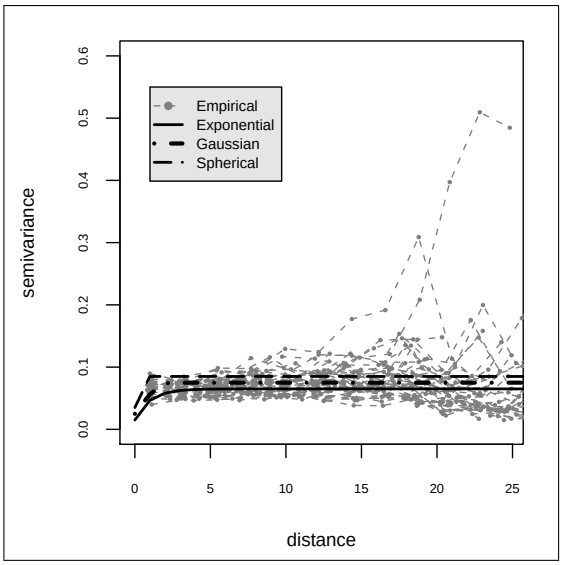
$$2\bar{\gamma}(\boldsymbol{h}) = \left\{ \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} |e(\boldsymbol{s}_i) - e(\boldsymbol{s}_j)|^{\frac{1}{2}} \right\}^4 \Bigg/ \left\{ 0.457 + \frac{0.494}{|N(\boldsymbol{h})|} \right\} \tag{1.5}$$

$N(\boldsymbol{h}) \equiv \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h} \ \forall \ i,j\}$ and $e(\boldsymbol{s}_i) = p(\boldsymbol{s}_i) - \boldsymbol{X}(\boldsymbol{s}_i)\hat{\boldsymbol{\beta}}$. This estimator is less sensitive to outliers and Cressie (1993) argues that it is more efficient than the classical estimator and other squared difference based estimators. It is also robust to some non-normality in the spatial process.

Inference requires that the variogram must be negative semi-definite. Since estimated (or empirical) variograms are not always negative semi-definite they cannot be

directly used for inference (Haining 2003, pg. 327). A theoretical variogram that approximates the empirical variogram and is negative semi-definite is used instead. Since the space of valid variograms is large, it is common practice to initially choose a parametric family of variograms and then choose the best fitting variogram from this family. The best-fitting variogram has spatial dependence that most closely approximates the dependence in the data. The empirical variograms from the 38 subsamples are plotted in Figure 1.3. Representative variograms from the exponential, spherical and Gaussian parametric variogram families are also plotted. All the parametric families fit the data. In the end the spatial covariance function $\rho(\cdot)$ is assumed to take the exponential

Fig. 1.3. Fitting Parametric Variograms to the Data



form, which is well behaved and used extensively in the literature.[4] For the exponential variogram, $\boldsymbol{\phi} = \{\phi\}$, called the range parameter.

To summarize the spatial model, $\ln(Price)$ is assumed to be a spatial process with trend $\boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta}$ and error $\epsilon(\boldsymbol{s})$. The error is zero-centered and distributed multivariate normal with $\epsilon(\boldsymbol{s}) \sim N(0, \sigma^2\boldsymbol{H}(\phi) + \tau^2\boldsymbol{I})$ where $[\boldsymbol{H}(\phi)]_{ij} = \rho(\phi; \boldsymbol{s}_i - \boldsymbol{s}_j)$ is ij^{th} element of the covariance matrix. Spatial covariance is assumed to decay exponentially since the exponential variogram fits the empirical variograms well.

1.3.2 Least Squares Estimators

From among the various Least Squares (LS) estimation methods the Iterated Weighted Least Squares (IWLS) method is chosen to estimate model parameters $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$. Cressie (1993, pg. 97) describes it as a "pragmatic compromise between efficiency (GLS) and simplicity (OLS)." GLS cannot be used because the correlation structure is unknown and OLS is infeasible because it does not permit spatial correlation. The IWLS method begins begins with the OLS estimation of $\boldsymbol{\beta}$. The residuals are calculated, and used to estimate the Hawkins-Cressie empirical variogram [see (1.5)]. LS estimates of $\boldsymbol{\alpha}$ are then obtained by substituting the empirical semi-variogram,[5] $\bar{\gamma}(\boldsymbol{h})$, into a squared errors-based loss function, which is then minimized. Two loss functions common to the literature are used, supplying two sets of Least Squares results. The first is referred to as the Bin-Weighted Loss function (1.6) and the second, which was proposed by Cressie

[4] Under an exponential distribution, $\rho(\phi; \boldsymbol{h}) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi\boldsymbol{h})) & \text{if } \boldsymbol{h} > \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases}$

[5] The semi-variogram, $\bar{\gamma}(\boldsymbol{h})$, is exactly half the variogram, $2\bar{\gamma}(\boldsymbol{h})$.

(1985), as the Cressie Loss function (1.7).

$$LOSS(\boldsymbol{\theta}) = \sum_j |N(\boldsymbol{h}(j))| \left\{\bar{\gamma}(\boldsymbol{h}(j)) - \gamma(\boldsymbol{h}(j);\boldsymbol{\theta})\right\}^2 \tag{1.6}$$

$$LOSS(\boldsymbol{\theta}) = \sum_j |N(\boldsymbol{h}(j))| \left\{\frac{\bar{\gamma}(\boldsymbol{h}(j))}{\gamma(\boldsymbol{h}(j);\boldsymbol{\theta})} - 1\right\}^2 \tag{1.7}$$

Minimizing (1.6) or (1.7) gives estimates of $\boldsymbol{\alpha} = \{\tau^2, \sigma^2, \phi\}$, called $\hat{\boldsymbol{\alpha}}_B$ or $\hat{\boldsymbol{\alpha}}_C$ respectively. The j's in (1.6) and (1.7) are the bins that the lags, $\boldsymbol{h}$, have been divided into. The fitting criterion in (1.7) gives more weight to observation pairs with smaller lags, which implies a good fit near the origin. As Stein (1988) notes, this is a good property. It also conforms to Tobler's First Law of Geography, which favors small distance relationships.

The GLS estimate of $\boldsymbol{\beta}$ is obtained by minimizing the expression in (1.8). However this expression is unknown because α_{GLS} is unknown.

$$(2\bar{\gamma}(\boldsymbol{h}) - 2\gamma(\boldsymbol{h};\boldsymbol{\theta}))' \boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1} (2\bar{\gamma}(\boldsymbol{h}) - 2\gamma(\boldsymbol{h};\boldsymbol{\theta})) \tag{1.8}$$

Instead, after minimizing (1.6) or (1.7) estimates $\hat{\boldsymbol{\beta}}_{OLS}$ and $\hat{\boldsymbol{\alpha}}_B$ or $\hat{\boldsymbol{\alpha}}_C$ are known. Suppressing subscripts for $\hat{\boldsymbol{\alpha}}$ let $\hat{\boldsymbol{\theta}}^{(0)} = \{\hat{\boldsymbol{\beta}}_{OLS}, \hat{\boldsymbol{\alpha}}\}$. $\hat{\boldsymbol{\theta}}^{(0)}$ is used to estimate the spatial covariance matrix $\boldsymbol{\Sigma}$ giving $\hat{\boldsymbol{\Sigma}}\big(\hat{\boldsymbol{\theta}}^{(0)}\big)$. This is substituted into equation (1.8), which when minimized provides $\hat{\boldsymbol{\theta}}^{(1)}$, a new estimate for $\boldsymbol{\theta}$. $\hat{\boldsymbol{\theta}}^{(1)}$ is used to estimate $\hat{\boldsymbol{\Sigma}}\big(\hat{\boldsymbol{\theta}}^{(1)}\big)$, which is substituted into (1.8), which when minimized gives $\hat{\boldsymbol{\theta}}^{(2)}$ and so on. The iterations continue until convergence and the calculation of IWLS estimates $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}\}$. Specifically,

as noted in Schabenberger and Gotway (2005, pg. 258)

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}})^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}})^{-1}\boldsymbol{p} \tag{1.9}$$

$$\hat{V}[\hat{\boldsymbol{\beta}}] = \left(\boldsymbol{X}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\alpha}})^{-1}\boldsymbol{X}\right)^{-1} \tag{1.10}$$

$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_B$ if the Bin-Weighted loss function (1.6) is used and $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_C$ if the Cressie loss function (1.7) is used.

1.3.3 Maximum Likelihood Estimators

Two Maximum Likelihood (ML) methods of estimation are examined: the classical ML estimator (M) and the Restricted ML estimator (R). Assuming that log of property price $\boldsymbol{p} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$, M estimates of $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ are obtained by minimizing the negative log-likelihood function of $\boldsymbol{p}$ in (1.11).

$$\mathfrak{L}(\boldsymbol{\theta}) = \frac{n}{2}\log 2\pi + \frac{1}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \frac{1}{2}(\boldsymbol{p} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{p} - \boldsymbol{X}\boldsymbol{\beta}) \tag{1.11}$$

The M estimator of the error parameters, $\hat{\boldsymbol{\alpha}}_M$, is severely biased (Mardia and Marshall 1984) because of the presence of $\boldsymbol{\beta}$ in the likelihood function. Patterson and Thompson (1971) devised the R estimator, which is independent of $\boldsymbol{\beta}$ and corrects the bias in the estimation of $\boldsymbol{\alpha}$. It works by applying the ML method to error contrasts rather than the data itself. An error contrast is a linear combination $\boldsymbol{a}'\boldsymbol{Q}$ with two properties: first $E(\boldsymbol{a}'\boldsymbol{Q}) = 0$ and second $\boldsymbol{a}'\boldsymbol{X} = \boldsymbol{0}'$. R estimates of $\boldsymbol{\theta}$ are obtained by minimizing the negative log-likelihood function in (1.12) where $\boldsymbol{\Pi}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$. By substituting $\hat{\boldsymbol{\theta}}_M$ and $\hat{\boldsymbol{\theta}}_R$ into (1.9) and (1.10),

M and R estimates of $\boldsymbol{\beta}$ and $V(\boldsymbol{\beta})$are obtained.

$$\mathfrak{L}(\boldsymbol{\theta}) = \frac{n-(p+1)}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{X}| + \frac{1}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}| + \frac{1}{2}\boldsymbol{Q}'\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{Q} \quad (1.12)$$

Prediction or kriging variances and the bias in error related estimates are lower under R when compared to M. M and R estimation is operationalized under many algorithms. In this study the the Nelder-Mead algorithm is used, chosen for its robustness properties (Nelder and Mead 1965).

1.3.4 Hierarchical Bayesian Estimators

Two levels of stochasticity are assumed under the Hierarchical Bayesian (H) estimation method. On the first level, $p(\boldsymbol{s})$ is assumed to be normally distributed, conditional on the covariate data $\boldsymbol{X}(\boldsymbol{s})$, model parameters $\boldsymbol{\theta}$ and a vector of spatial effects $\boldsymbol{W}(\boldsymbol{s})$. On the second level, the spatial effects $\boldsymbol{W}(\boldsymbol{s})$ are assumed to be normally distributed, conditional on $\boldsymbol{\theta}$, which is also stochastic. Note that in the Bayesian method, model parameters are treated as random effects, whereas in the Classical approach, model parameters are fixed.

Estimation under the H method is as follows: first, $\boldsymbol{\theta}$ is assigned a prior distribution. Next data is used to confront the prior and update beliefs about the behavior of $\boldsymbol{\theta}$. The updated beliefs take the form of a posterior distribution for $\boldsymbol{\theta}$. The process of moving from prior to posterior beliefs is underpinned by Bayes' Theorem. The posterior

distribution of $\boldsymbol{\theta}$ conditional on $\boldsymbol{p}$ and $\boldsymbol{X}$, $\pi(\boldsymbol{\theta}|\boldsymbol{p},\boldsymbol{X})$ is

$$\pi(\boldsymbol{\theta}|\boldsymbol{p},\boldsymbol{X}) = \frac{f(\boldsymbol{p}|\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{X})f(\boldsymbol{W}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\boldsymbol{p}|\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{X})f(\boldsymbol{W}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{1.13}$$

where $f(\boldsymbol{p}|\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{X})$ is the prior distribution assumed for $\boldsymbol{p}$ conditional on $\boldsymbol{\theta}$, $\boldsymbol{W}$ and $\boldsymbol{X}$, $f(\boldsymbol{W}|\boldsymbol{\theta})$ is the prior on $\boldsymbol{W}$ conditional on $\boldsymbol{\theta}$, and $f(\boldsymbol{\theta})$ is the prior on $\boldsymbol{\theta}$. *A priori* it is assumed that the parameters are independent: $f(\boldsymbol{\theta}) = f(\boldsymbol{\beta})f(\sigma^2)f(\tau^2)f(\phi)$.

From (1.2), (1.3) and (1.4), $\boldsymbol{p}|\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{W},\tau^2\boldsymbol{I})$ and $\boldsymbol{W}|\boldsymbol{\theta} \sim N(\boldsymbol{0},\sigma^2\boldsymbol{H}(\phi))$. σ^2 and τ^2 are assigned inverse gamma priors,[6] $\boldsymbol{\beta}$ has a flat prior[7] and ϕ is distributed uniformly.[8] These distributional assumptions are common in the geostatistical literature (Banerjee et al. 2004, pg. 131) because they are flexible enough to approximate a variety of spatial processes. The flat prior for the $\boldsymbol{\beta}$s implies that the H method is asymptotically equivalent to the M method (Banerjee et al. 2004, pg. 103).

From the posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\boldsymbol{p},\boldsymbol{X})$, the posterior distributions of the individual parameters are calculated by marginalization. For example the posterior distribution of ϕ is

$$\pi(\phi|\boldsymbol{p},\boldsymbol{X}) = \frac{\int\int\int f(\boldsymbol{p}|\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{X})f(\boldsymbol{W}|\boldsymbol{\theta})f(\boldsymbol{\beta})f(\tau^2)f(\sigma^2)f(\phi)\ d\boldsymbol{\beta}\ d\tau^2\ d\sigma^2}{\int f(\boldsymbol{p}|\boldsymbol{\theta},\boldsymbol{W},\boldsymbol{X})f(\boldsymbol{W}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{1.14}$$

---

[6] The inverse gamma distribution is defined over the support $x \in (0,\infty)$ and has the *pdf* $f(x;a,b) = \frac{b^a}{\Gamma(a)}x^{-a-1}\exp\left(\frac{-b}{x}\right)$.

[7] $\boldsymbol{\beta} \sim N(\bar{\boldsymbol{\beta}}, diag(\infty))$. The infinite variance of each element of $\boldsymbol{\beta}$ implies that $\boldsymbol{\beta}$ has zero mass along its entire range. $\boldsymbol{\beta}$ is essentially left undefined.

[8] $\phi \sim U(a,b)$, where a and b define the range.

The posterior distributions of other parameters are analogously calculated. Once the marginal posterior distributions are known, parameter statistics are easily calculated. For example the posterior mean and posterior median of ϕ are calculated as shown in (1.15) and (1.16).

$$\text{Posterior Mean:} \quad \bar{\phi} = \int \phi \, d\pi(\phi|\boldsymbol{p}, \boldsymbol{X}) \tag{1.15}$$

$$\text{Posterior Median:} \quad \hat{\phi} = \int_{-\infty}^{\hat{\phi}} d\pi(\phi|\boldsymbol{p}, \boldsymbol{X}) = 0.5 \tag{1.16}$$

Often $\pi(\boldsymbol{\theta}|\boldsymbol{p}, \boldsymbol{X})$ and the marginal posteriors are not available in closed form, or the closed form solutions are so messy that calculating statistics becomes intractable. Then MCMC methods are used to sample from the posteriors. A correlated sample is collected by iteratively drawing from $f(\boldsymbol{p}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X}) f(\boldsymbol{W}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$, which is proportional to $\pi(\boldsymbol{\theta}|\boldsymbol{p}, \boldsymbol{X})$ [see (1.13)] since the denominator is constant. Draws from $\pi(\boldsymbol{\theta}|\boldsymbol{p}, \boldsymbol{X})$ are not made because calculating the denominator is computationally expensive. The correlated sample is then used to estimate desired statistics. Increasing the length of the chain increases estimator accuracy.

A number of algorithms can be used to operationalize MCMC sampling. In this study the Gibbs sampler is used to draw the $\boldsymbol{\beta}$ sample and the Metropolis-Hastings algorithm to draw the $\boldsymbol{\alpha}$ sample. The Gibbs sampler is easier to implement, but can only be used when the full conditional distribution $\pi(\beta_i|\boldsymbol{\beta}_{-i}, \sigma^2, \tau^2, \phi, \boldsymbol{p}, \boldsymbol{X})$ is available in closed form; which occurs only when the prior $f(\boldsymbol{\beta})$ and the likelihood $f(\boldsymbol{p}|\boldsymbol{\theta})$ are a conjugate pair (Banerjee et al. 2004, pg. 113). Since closed forms of the full conditionals of $\boldsymbol{\alpha}$ are not available, the Metropolis-Hastings algorithm is used to obtain samples of $\boldsymbol{\alpha}$.

A correlated sample is drawn because then the M-H algorithm converges quicker. The quality of the chains are ensured through diagnostics that checked the suitability of the priors and the convergence of the chain. The appropriateness of the priors was analyzed through the acceptance rate. For all estimations the acceptance rates were in the 25-40% range, which is optimal (Gelman et al. 1996). Chain convergence was checked through tests devised by Geweke (1992) and Heidelberger and Welch (1983). The tests indicated that all MCMC simulations chosen for analysis did converge.

1.3.5 Prediction or Kriging

Prediction in a spatial setting is commonly called kriging. In spatial analysis, it is often important to generate good predictions at locations in the geographical window. Analysis of the predictive capabilities of the estimation methods supplies a powerful signal vis-á-vis the desirability of the approaches. But first I present some theory about kriging under Classical and Bayesian approaches. Let $\boldsymbol{s}_0$ be a vector of m locations where prediction is desired and $\boldsymbol{s}_1$ be the vector of the n locations in the sample. Subsuming the location argument $\boldsymbol{s}$ and since p is normally distributed

$$\begin{pmatrix} \boldsymbol{p}_0 \\ \boldsymbol{p}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{X}_0\boldsymbol{\beta} \\ \boldsymbol{X}_1\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_0 \\ \boldsymbol{\Sigma}_0^T & \boldsymbol{\Sigma} \end{pmatrix} \right) \tag{1.17}$$

where $\boldsymbol{\Sigma}$ is obtained from (1.4), $\boldsymbol{\Sigma}_{00}$ is an $m \times m$ matrix of covariances across predictee locations and $\boldsymbol{\Sigma}_0$ an $m \times n$ matrix of covariances between predictee and observed locations, all matrices defined by the spatial covariance function $\rho(\cdot)$. The normality assumption supplies the closed form for the conditional distribution $\boldsymbol{p}_0|\boldsymbol{p}_1$, the mean and variance of

which are

$$E[\boldsymbol{p}_0|\boldsymbol{p}_1] = \boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}(\boldsymbol{p}_1 - \boldsymbol{X}\boldsymbol{\beta}) \tag{1.18}$$

$$V[\boldsymbol{p}_0|\boldsymbol{p}_1] = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0' + \left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)' \tag{1.19}$$

The mean (1.18) is also called the Best Linear Unbiased Predictor (BLUP) under squared-error loss (Gotway and Cressie 1993). To get Classical estimates of the kriged means and variances at $\boldsymbol{s}_0$ conditional on data at $\boldsymbol{s}_1$ the Classical parameter estimates ($\hat{\boldsymbol{\theta}}_B$ or $\hat{\boldsymbol{\theta}}_C$ or $\hat{\boldsymbol{\theta}}_M$ or $\hat{\boldsymbol{\theta}}_R$ in this paper) are substituted into (1.18) and (1.19). When kriging in the Hierarchical framework, the predictive distribution of the spatial process $p(\boldsymbol{s})$ at the predictee locations $\boldsymbol{s}_0$ is

$$f(\boldsymbol{p}_0|\boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{X}_1) = \int f(\boldsymbol{p}_0|\boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{p}_1, \boldsymbol{X}_1)\, d\boldsymbol{\theta} \tag{1.20}$$

$f(\boldsymbol{p}_0|\boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{\theta})$ is conditionally normal and generates (1.18) and (1.19) through integration. $\pi(\boldsymbol{\theta}|\boldsymbol{p}_1, \boldsymbol{X}_1)$ is known from (1.13). Draws proportional to $\pi(\boldsymbol{\theta}|\boldsymbol{p}_1, \boldsymbol{X}_1)$ already exist because of the MCMC methods used to estimate $\hat{\beta}_H$ and $\hat{\alpha}_H$. These draws are used to generate draws from the distribution of $\boldsymbol{p}_0|\boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{\theta}$, which is known to be normally distributed. The latter draws are then used to generate density plots or to estimate statistics related to $\boldsymbol{p}_0$.

1.4 Results and Discussion

Inference on the performance of the methods of estimation is based on the rankings of the model parameter estimators $\hat{\boldsymbol{\theta}}$ and price prediction estimators $\hat{\boldsymbol{p}}_0$ from each method. The estimators are ranked on the basis of their accuracy, precision and overall quality. Accuracy is measured through estimator bias or distance between estimate and true value. Precision is determined from the spread or variability of the estimates: precise estimators are less variable. Quality is gauged through metrics that combine information on accuracy and precision.

1.4.1 Estimating the Trend Component

For any estimator of the trend coefficient $\hat{\beta}$, bias is $b(\beta, \hat{\beta}) = E(\hat{\beta}) - \beta$ where $E(\hat{\beta})$ is the expected value of the estimator $\hat{\beta}$ and $\beta \in \boldsymbol{\beta}$ is the true value of the trend coefficient. The true value β is unknown, which precludes an exact measurement of bias and exact inference about accuracy. However, as discussed in Section 1.1, I do have *a priori* expectations about the sign of β. By comparing the sign of $\hat{\beta}$ to the *a priori* expectation, some inference about accuracy is possible. For example, if a positive sign for β is expected and the estimate from one method is positive while the other negative, one may infer that the first method provides a more accurate estimator of β. If the estimates from both methods are positive and not significantly different one may infer that both methods supply equally accurate estimators of β. Not knowing the true value of β prevents inference only when both estimates are positive *and* significantly different.

From the regressions of the subsamples, for each estimation method 38 estimates of each β are obtained. The plug-in estimator of $E(\hat{\beta})$ is $\bar{\beta} = \sum_{k=1}^{38} \hat{\beta}_k/k$ where $\hat{\beta}_k$ is the kth partial jackknife estimate of $\beta \in \boldsymbol{\beta}$. Comparison of the $\bar{\beta}$s obtained from the different methods supplies inference about the mean accuracy of these methods with respect to estimation of the trend component. Values of $\bar{\beta}$ corresponding to each variable in the trend, across the five methods of estimation are the unbracketed values in Table 1.2. All $\bar{\beta}$s without superscripts are significantly different from zero at the 1% level. The † superscript implies significance at 5% level. The * superscript indicates that the coefficient is not significantly different from zero at the 5% level.

1.4.2 Estimating the Trend Component

For any estimator of the trend coefficient $\hat{\beta}$, bias is $b(\beta, \hat{\beta}) = E(\hat{\beta}) - \beta$ where $E(\hat{\beta})$ is the expected value of the estimator $\hat{\beta}$ and $\beta \in \boldsymbol{\beta}$ is the true value of the trend coefficient. The true value β is unknown, which precludes an exact measurement of bias and exact inference about accuracy. However, as discussed in Section 1.1, I do have *a priori* expectations about the sign of β. By comparing the sign of $\hat{\beta}$ to the *a priori* expectation, some inference about accuracy is possible. For example, if a positive sign for β is expected and the estimate from one method is positive while the other negative, one may infer that the first method provides a more accurate estimator of β. If the estimates from both methods are positive and not significantly different one may infer that both methods supply equally accurate estimators of β. Not knowing the true value of β prevents inference only when both estimates are positive *and* significantly different.

From the regressions of the subsamples, for each estimation method 38 estimates of each β are obtained. The plug-in estimator of $E(\hat{\beta})$ is $\bar{\beta} = \sum_{k=1}^{38} \hat{\beta}_k/k$ where $\hat{\beta}_k$ is the kth partial jackknife estimate of $\beta \in \boldsymbol{\beta}$. Comparison of the $\bar{\beta}$s obtained from the different methods supplies inference about the mean accuracy of these methods with respect to estimation of the trend component. Values of $\bar{\beta}$ corresponding to each variable in the trend, across the five methods of estimation are the unbracketed values in Table 1.2. All $\bar{\beta}$s without superscripts are significantly different from zero at the 1% level. The † superscript implies significance at 5% level. The * superscript indicates that the coefficient is not significantly different from zero at the 5% level.

Table 1.2: Metrics for Accuracy of $\hat{\boldsymbol{\theta}}$

Variable	B	C	M	R	H
$Intercept$	16.758	16.588	16.547	16.569	17.031
	(100.00)	(100.00)	(100.00)	(100.00)	(100.00)
$Northing$	0.030	0.033	0.028	0.031	0.034
	(100.00)	(100.00)	(100.00)	(97.37)	(97.37)
$Easting$	$-0.011^{\dagger}$	-0.011^{*}	-0.015	-0.013	-0.013
	(34.21)	(36.84)	(31.58)	(31.58)	(34.21)
$\#Baths$	0.086	0.083	0.088	0.086	0.083
	(89.47)	(86.84)	(92.11)	(97.37)	(97.37)
$\#Garages$	0.030	0.026	0.033	0.030	0.026
	(78.95)	(73.68)	(86.84)	(78.95)	(73.68)
$\ln(Area)$	0.633	0.617	0.658	0.651	0.638
	(100.00)	(100.00)	(100.00)	(100.00)	(100.00)
$Carpet$	0.015^{*}	0.011^{*}	0.036	0.037	0.034
	(60.53)	(63.16)	(71.05)	(71.05)	(68.42)
$DiningRoom$	0.041	0.036	0.049	0.043	0.032
	(78.95)	(71.05)	(73.68)	(73.68)	(63.16)
$24DoorKeep$	0.067	0.063	0.080	0.078	0.072
	(76.32)	(73.68)	(92.11)	(86.84)	(76.32)
$\ln(PM_{10})$	-0.296	-0.349	-0.286	-0.306	-0.316
	(84.21)	(81.58)	(78.95)	(78.95)	(78.95)
$\ln(ZonalPark)$	0.031	0.024	0.037	0.035	0.027
	(23.68)	(23.68)	(18.42)	(23.68)	(23.68)
$\ln(FloodArea)$	$0.019^{\dagger}$	0.033	$0.019^{\dagger}$	0.026	0.028
	(60.53)	(65.79)	(63.16)	(68.42)	(65.79)
$\ln(72^{nd}St)$	-0.052	-0.006^{*}	-0.055	-0.043	-0.060
	(76.32)	(60.53)	(73.68)	(68.42)	(65.79)

Variable	B	C	M	R	H
$\ln(HistCent)$	-0.166	-0.170	-0.161	-0.170	-0.181
	(86.84)	(78.95)	(84.21)	(84.21)	(84.21)
$CrimeIndex$	-0.001^{*}	-0.002^{*}	-0.001^{*}	-0.001^{*}	-0.001^{*}
	(47.37)	(55.26)	(55.26)	(57.89)	(52.63)
$Stratum$	0.223	0.200	0.235	0.225	0.207
	(100.00)	(100.00)	(100.00)	(100.00)	(100.00)
Nugget, τ^2	$0.003^{\dagger}$	0.032	0.025	0.033	0.035
Partial Sill, σ^2	0.072	6.451^{*}	0.052	0.338	0.108
Range, ϕ	1.331	$1399.801^{\dagger}$	0.272	$18.684^{\dagger}$	0.477

† and * indicate significance and nonsignificance of mean estimates at the 5% level. All other mean estimates are significant at the 1% level.

The $\bar{\beta}$s corresponding to all independent variables are identical in sign and similar in magnitude across the five methods of estimation. The $\bar{\beta}$s also have the *a priori* expected sign for all independent variables except $Easting$ and $\ln(ZonalPark)$. $\bar{\beta}_{Easting}$ was expected to be positive on the basis of the contour map of $\ln p$ in Figure 1.1(b), which showed a broad upward trend in the northeast direction. However, the map also reveals pockets of high priced neighborhoods scattered throughout north Bogotá. Some subsamples include high value properties located west of lower value eastern properties, which results in a negative sign for $\bar{\beta}_{Easting}$. Low $\bar{\beta}_{Easting}$ values across the five methods implies that relocation along the East-West axis has small effect on property sale price. The positive sign on $\bar{\beta}_{\ln(ZonalPark)}$ is explained by anecdotal evidence that zonal parks are poorly maintained and often used by gangs for drug and other criminal transactions. With *a priori* expectation mostly met the $\bar{\beta}$ coefficients show that the Classical and Bayesian methods of estimation are equally accurate at estimating the vector of trend coefficients $\boldsymbol{\beta}$ in a spatial hedonic property model: $M \sim H \sim B \sim R \sim C$.

Inference with regard to the accuracy of the $\hat{\beta}$s is also possible through comparison of the proportion of $\hat{\beta}_k$s with the expected sign for each estimation procedure. One may

infer that the most accurate method has the greatest proportion of $\hat{\beta}_k$s with the correct sign. Looking at $\bar{\beta}$ allowed inference about accuracy in the mean. Looking at proportions on the other hand allows inference about the likelihood that the estimate $\hat{\beta}$ will have the expected sign. The proportions are the bracketed numbers in Table 1.2. Some $\hat{\beta}$s take the sign contrary to expectation in a high proportion of the estimations. $\hat{\beta}_{Easting}$ and $\hat{\beta}_{\ln(ZonalPark)}$ have the expected sign in less than 40% of the regressions, which is unsurprising given that $\bar{\beta}_{Easting}$ and $\bar{\beta}_{\ln(ZonalPark)}$ also had the wrong sign. $\hat{\beta}_{CrimeIndex}$ has the expected sign in about 50% of the regressions across methods, implying unskewed deviations about zero. This result is also expected because, for all methods, $CrimeIndex$ is the only independent variable that does not significantly affect $\ln p$.

Not counting $\hat{\beta}_{East}$, $\hat{\beta}_{Zone}$ and $\hat{\beta}_{Crime}$ the mean proportion of $\hat{\beta}$s with the correct sign is calculated for each estimation method by averaging down the columns in Table 1.2. The mean proportion of $\hat{\beta}$s with the correct sign is over 80% for all methods. The C method does the worst with 81.17% of the $\hat{\beta}^C$s having the expected sign and the M method does best with 85.83% of the $\hat{\beta}^M$s having the expected sign. Paired t-tests[9] indicate that the differences in mean proportion are not significant at the 10% level, which implies that the proportion of $\hat{\beta}$s with the expected sign are statistically similar across the five methods: $M \sim H \sim R \sim B \sim C$. Supporting the $\bar{\beta}$ results analysis of proportions shows that the Classical and Bayesian methods estimate the vector of trend coefficients $\boldsymbol{\beta}$ in a spatial hedonic property model equally accurately.

[9]The paired t-test tests if two $\bar{\beta}$s, say $\bar{\beta}_M$ and $\bar{\beta}_C$ are significantly different. The test statistic is $t_{M,C} = (\bar{\beta}_M - \bar{\beta}_C)/\sqrt{(se(\bar{\beta}_M) + se(\bar{\beta}_C))/38}$.

Inference about the precision of the different methods when estimating $\hat{\beta}$ is drawn from analysis of the dispersion of $\hat{\beta}$ since, in the statistical sense, precision is defined as the inverse of the variance (Banerjee et al. 2004, pg. 101). Since standard deviation of $\hat{\beta}$, $sd(\hat{\beta}) = \sqrt{V(\hat{\beta})} = \sqrt{1/\psi(\hat{\beta})}$ where $\psi(\cdot)$ measures precision, analysis of sd allows inference about precision. The lower the sd the greater the precision. The unbracketed numbers in Table 1.4 are the standard deviations of trend and error coefficients across the five compared methods of estimation.

Table 1.3: Standard Deviation and Range of $\hat{\theta} = \{\hat{\beta}, \hat{\alpha}\}$

Variable	B	C	M	R	H
$Intercept$	3.029	3.858	2.879	3.478	3.223
	(11.87)	(15.12)	(11.29)	(13.63)	(12.63)
$Northing$	0.018	0.027	0.018	0.023	0.021
	(0.070)	(0.105)	(0.070)	(0.092)	(0.082)
$Easting$	0.030	0.036	0.026	0.030	0.029
	(0.119)	(0.139)	(0.101)	(0.116)	(0.114)
$\#Baths$	0.065	0.067	0.055	0.053	0.052
	(0.253)	(0.262)	(0.215)	(0.208)	(0.205)
$\#Garages$	0.041	0.042	0.039	0.041	0.038
	(0.160)	(0.165)	(0.154)	(0.160)	(0.148)
$\ln(Area)$	0.118	0.132	0.091	0.095	0.090
	(0.461)	(0.519)	(0.356)	(0.371)	(0.351)
$Carpet$	0.086	0.092	0.081	0.078	0.084
	(0.338)	(0.360)	(0.318)	(0.307)	(0.328)
$DiningRoom$	0.068	0.069	0.076	0.074	0.073
	(0.267)	(0.271)	(0.300)	(0.290)	(0.288)
$24DoorKeep$	0.085	0.090	0.060	0.063	0.070
	(0.335)	(0.354)	(0.236)	(0.246)	(0.275)
$\ln(PM_{10})$	0.308	0.434	0.314	0.334	0.337
	(1.209)	(1.703)	(1.230)	(1.308)	(1.321)
$\ln(ZonalPark)$	0.047	0.053	0.045	0.044	0.046
	(0.185)	(0.209)	(0.176)	(0.174)	(0.181)
$\ln(FloodArea)$	0.051	0.067	0.047	0.052	0.058
	(0.201)	(0.264)	(0.186)	(0.203)	(0.228)
$\ln(72^{nd}St)$	0.102	0.145	0.089	0.097	0.108
	(0.398)	(0.567)	(0.348)	(0.382)	(0.424)
$\ln(HistCent)$	0.193	0.257	0.192	0.235	0.230
	(0.757)	(1.006)	(0.754)	(0.923)	(0.900)
$CrimeIndex$	0.009	0.010	0.009	0.009	0.010
	(0.034)	(0.037)	(0.035)	(0.034)	(0.038)

Variable	B	C	M	R	H
Stratum	0.064	0.076	0.055	0.059	0.057
	(0.250)	(0.298)	(0.214)	(0.230)	(0.222)
Nugget, τ^2	0.010	0.026	0.029	0.029	0.005
	(0.040)	(0.101)	(0.114)	(0.114)	(0.021)
Partial Sill, σ^2	0.019	22.055	0.043	0.750	0.009
	(0.074)	(86.45)	(0.168)	(2.942)	(0.033)
Range, ϕ	3.061	4070.194	0.377	54.875	0.012
	(12.00)	(15955.16)	(1.479)	(215.11)	(0.048)

The standard deviation of $\hat{\beta}^H_{\ln(Area)}$ is lowest followed by, in ascending order, $\hat{\beta}^M_{\ln(Area)}$, $\hat{\beta}^R_{\ln(Area)}$, $\hat{\beta}^B_{\ln(Area)}$ and $\hat{\beta}^C_{\ln(Area)}$. By likening the *sd*s to scores for each method, the results indicate that H method supplies the most precise estimates of $\ln(Area)$ and the C method supplies the least precise estimates. Thinking of the *sd*s associated with each of the 16 independent variables as scores implies a set of 16 precision scores for each estimation method. Inference on any pair of methods may then drawn by comparing their scores through tests like the nonparametric Wilcoxon Signed Rank test. This test is commonly used for the rank-based comparison of two matched samples because it has a high power efficiency when compared to similar tests (Hays and Winkler 1971). Given five methods of estimation there are ten possible pairings. The Wilcoxon test is applied to each pair to determine if the precision of one is significantly greater than the other. Significance is determined at the 5% level. The tests show that the precision of the M method is significantly greater than the precision of all other methods. The precision of the C method is significantly worse than the precision of the other methods. The precision of the B, R and H methods do not differ significantly. One may infer that the M method is the most precise in the estimation of $\hat{\boldsymbol{\beta}}$, the C method is the least precise. The other methods lie between these extremes and perform equivalently. In order of precision the methods may be ranked $M \succ B \sim R \sim H \succ C$.

Inference with regard to the accuracy of the $\hat{\beta}$s is also possible through comparison of the proportion of $\hat{\beta}_k$s with the expected sign for each estimation procedure. One may infer that the most accurate method has the greatest proportion of $\hat{\beta}_k$s with the correct sign. Looking at $\bar{\beta}$ allowed inference about accuracy in the mean. Looking at proportions on the other hand allows inference about the likelihood that the estimate $\hat{\beta}$ will have the expected sign. The proportions are the bracketed numbers in Table 1.2. Some $\hat{\beta}$s take the sign contrary to expectation in a high proportion of the estimations. $\hat{\beta}_{Easting}$ and $\hat{\beta}_{\ln(ZonalPark)}$ have the expected sign in less than 40% of the regressions, which is unsurprising given that $\bar{\beta}_{Easting}$ and $\bar{\beta}_{\ln(ZonalPark)}$ also had the wrong sign. $\hat{\beta}_{CrimeIndex}$ has the expected sign in about 50% of the regressions across methods, implying unskewed deviations about zero. This result is also expected because, for all methods, $CrimeIndex$ is the only independent variable that does not significantly affect $\ln p$.

Not counting $\hat{\beta}_{East}$, $\hat{\beta}_{Zone}$ and $\hat{\beta}_{Crime}$ the mean proportion of $\hat{\beta}$s with the correct sign is calculated for each estimation method by averaging down the columns in Table 1.2. The mean proportion of $\hat{\beta}$s with the correct sign is over 80% for all methods. The C method does the worst with 81.17% of the $\hat{\beta}^C$s having the expected sign and the M method does best with 85.83% of the $\hat{\beta}^M$s having the expected sign. Paired t-tests[10] indicate that the differences in mean proportion are not significant at the 10% level, which implies that the proportion of $\hat{\beta}$s with the expected sign are statistically similar across the five methods: $M \sim H \sim R \sim B \sim C$. Supporting the $\bar{\beta}$ results analysis of

[10]The paired t-test tests if two $\bar{\beta}$s, say $\bar{\beta}_M$ and $\bar{\beta}_C$ are significantly different. The test statistic is $t_{M,C} = (\bar{\beta}_M - \bar{\beta}_C)/\sqrt{(se(\bar{\beta}_M) + se(\bar{\beta}_C))/38}$.

proportions shows that the Classical and Bayesian methods estimate the vector of trend coefficients $\boldsymbol{\beta}$ of a spatial hedonic property model equally accurately.

Inference about the precision of the different methods when estimating $\hat{\beta}$ is drawn from analysis of the dispersion of $\hat{\beta}$ since, in the statistical sense, precision is defined as the inverse of the variance (Banerjee et al. 2004, pg. 101). Since standard deviation of $\hat{\beta}$, $sd(\hat{\beta}) = \sqrt{V(\hat{\beta})} = \sqrt{1/\psi(\hat{\beta})}$ where $\psi(\cdot)$ measures precision, analysis of sd allows inference about precision. The lower the sd the greater the precision. The unbracketed numbers in Table 1.4 are the standard deviations of trend and error coefficients across the five compared methods of estimation.

Table 1.4: Standard Deviation and Range of $\hat{\theta} = \{\hat{\beta}, \hat{\alpha}\}$

Variable	B	C	M	R	H
$Intercept$	3.029	3.858	2.879	3.478	3.223
	(11.87)	(15.12)	(11.29)	(13.63)	(12.63)
$Northing$	0.018	0.027	0.018	0.023	0.021
	(0.070)	(0.105)	(0.070)	(0.092)	(0.082)
$Easting$	0.030	0.036	0.026	0.030	0.029
	(0.119)	(0.139)	(0.101)	(0.116)	(0.114)
$\#Baths$	0.065	0.067	0.055	0.053	0.052
	(0.253)	(0.262)	(0.215)	(0.208)	(0.205)
$\#Garages$	0.041	0.042	0.039	0.041	0.038
	(0.160)	(0.165)	(0.154)	(0.160)	(0.148)
$\ln(Area)$	0.118	0.132	0.091	0.095	0.090
	(0.461)	(0.519)	(0.356)	(0.371)	(0.351)
$Carpet$	0.086	0.092	0.081	0.078	0.084
	(0.338)	(0.360)	(0.318)	(0.307)	(0.328)
$DiningRoom$	0.068	0.069	0.076	0.074	0.073
	(0.267)	(0.271)	(0.300)	(0.290)	(0.288)
$24DoorKeep$	0.085	0.090	0.060	0.063	0.070
	(0.335)	(0.354)	(0.236)	(0.246)	(0.275)
$\ln(PM_{10})$	0.308	0.434	0.314	0.334	0.337
	(1.209)	(1.703)	(1.230)	(1.308)	(1.321)
$\ln(ZonalPark)$	0.047	0.053	0.045	0.044	0.046
	(0.185)	(0.209)	(0.176)	(0.174)	(0.181)
$\ln(FloodArea)$	0.051	0.067	0.047	0.052	0.058
	(0.201)	(0.264)	(0.186)	(0.203)	(0.228)
$\ln(72^{nd}St)$	0.102	0.145	0.089	0.097	0.108
	(0.398)	(0.567)	(0.348)	(0.382)	(0.424)

Variable	B	C	M	R	H
$\ln(HistCent)$	0.193	0.257	0.192	0.235	0.230
	(0.757)	(1.006)	(0.754)	(0.923)	(0.900)
$CrimeIndex$	0.009	0.010	0.009	0.009	0.010
	(0.034)	(0.037)	(0.035)	(0.034)	(0.038)
$Stratum$	0.064	0.076	0.055	0.059	0.057
	(0.250)	(0.298)	(0.214)	(0.230)	(0.222)
Nugget, τ^2	0.010	0.026	0.029	0.029	0.005
	(0.040)	(0.101)	(0.114)	(0.114)	(0.021)
Partial Sill, σ^2	0.019	22.055	0.043	0.750	0.009
	(0.074)	(86.45)	(0.168)	(2.942)	(0.033)
Range, ϕ	3.061	4070.194	0.377	54.875	0.012
	(12.00)	(15955.16)	(1.479)	(215.11)	(0.048)

The standard deviation of $\hat{\beta}^H_{\ln(Area)}$ is lowest followed by, in ascending order, $\hat{\beta}^M_{\ln(Area)}$, $\hat{\beta}^R_{\ln(Area)}$, $\hat{\beta}^B_{\ln(Area)}$ and $\hat{\beta}^C_{\ln(Area)}$. By likening the *sd*s to scores for each method, the results indicate that H method supplies the most precise estimates of $\ln(Area)$ and the C method supplies the least precise estimates. Thinking of the *sd*s associated with each of the 16 independent variables as scores implies a set of 16 precision scores for each estimation method. Inference on any pair of methods may then drawn by comparing their scores through tests like the nonparametric Wilcoxon Signed Rank test. This test is commonly used for the rank-based comparison of two matched samples because it has a high power efficiency when compared to similar tests (Hays and Winkler 1971). Given five methods of estimation there are ten possible pairings. The Wilcoxon test is applied to each pair to determine if the precision of one is significantly greater than the other. Significance is determined at the 5% level. The tests show that the precision of the M method is significantly greater than the precision of all other methods. The precision of the C method is significantly worse than the precision of the other methods. The precision of the B, R and H methods do not differ significantly. One may infer that the M method is the most precise in the estimation of $\hat{\boldsymbol{\beta}}$, the C method is the least precise.

The other methods lie between these extremes and perform equivalently. In order of precision the methods may be ranked $M \succ B \sim R \sim H \succ C$.

Inference about relative precision may also be drawn by comparing the range of every $\hat{\beta}$ estimated under the different methods. Range is defined as $r(\hat{\beta}) = \max(\hat{\beta}) - \min(\hat{\beta})$. The ranges for every estimator of $\beta \in \boldsymbol{\beta}$ are reported as the bracketed values in Table 1.4. A smaller range implies greater clustering of the $\hat{\beta}$ estimator and hence greater precision. As with the $sd(\hat{\beta})$ statistics, $r(\hat{\beta})$ statistics may also be interpreted as scores. The Wilcoxon Signed Rank test is used to compare the range scores of the different methods. Methods are tested pair-wise to determine if the range of one is significantly greater than the other. Significance is determined at the 5% level. The test results support the *sd* comparison results: the M method has the significantly smallest range, the C method has the significantly largest range and the other methods lie between these extremes and do not differ significantly from each other. The methods may be ranked $M \succ B \sim R \sim H \succ C$ in order of precision.

For the special case when $\hat{\beta} = \bar{\beta}$ it is easy to prove that $MSE(\bar{\beta}) = V(\bar{\beta})$ since $b(\bar{\beta}, \beta) = 0$. Also note that $V(\bar{\beta}) = sd(\bar{\beta})^2 = sd(\hat{\beta})^2/k$ where $\bar{\beta}$ is the mean of $\hat{\beta}$. These results imply that $MSE(\bar{\theta})$ is a monotonic transformation of $sd(\hat{\theta})$, which in turn implies that the inferences drawn on the precision of $\hat{\beta}$ are applicable to the quality of these estimators in the mean. By induction the M method supplies estimators with the lowest MSE and the C method supplies estimators with the highest MSE. The MSEs of the other methods lie between these extremes and do not differ significantly from each other. One may infer that the M method supplies estimators of the highest quality and the C method supplies estimators of the lowest quality. If the quality of $\boldsymbol{\beta}$

estimators is desirable then the preference ordering of the five methods of estimation is $M \succ B \sim R \sim H \succ C$.

1.4.3 Estimating the Error Component

Since all the error coefficient estimates $\hat{\boldsymbol{\alpha}}$ are positive one cannot draw inference about accuracy by studying patterns in sign. The methods used to infer the accuracy of the $\hat{\beta}$s cannot be used to infer the accuracy of the $\hat{\alpha}$s. Instead I draw inference by individually considering the estimates of $\boldsymbol{\alpha}$ across methods. The plug-in estimator of $E(\hat{\alpha})$ is $\bar{\alpha} = \sum \hat{\alpha}_k / k$ where $\hat{\alpha}_k$ is the kth partial jackknife estimate of $\alpha \in \boldsymbol{\alpha} = \{\tau^2, \sigma^2, \phi\}$. Values of $\bar{\alpha}$ are reported in Table 1.2.

The range parameter ϕ defines the distance over which spatial correlation effects may be discerned. The effective range of spatial dependence, which is the distance at which the correlation drops to 0.05, is about $-\log(0.05)/\phi$ (Finley et al. 2007). The mean effective ranges are $2.25km$ for the B method, $2m$ for the C method, $11.03km$ for the M method, $161m$ for the R method and $6.29km$ for the H method. An effective range of $2m$ or even $161m$ implies an approximately non-spatial model given the geographical scale of the Bogotá property market. The empirical variogram in Figure 1.3 on the other hand implies the existence significant spatial effects. One may infer that the C and R methods supply biased and inaccurate range estimates and that the C method is the more inaccurate of the two. The mean ranges estimated by the other methods are more plausible. I infer that the B, M and H methods supply reasonably accurate estimates of ϕ, but further inference regarding the relative accuracy of these methods is not possible.

The nugget τ^2 captures idiosyncratic and microscale variations. For the C, M, R and H procedures $\bar{\tau}^2 \in [0.025, 0.035]$ while $\bar{\tau}^2_B = 0.003$. The empirical distribution of shows that $\hat{\tau}^2_B$ is bi-modal, with $\hat{\tau}^2 = 0$ in 90% of the B estimations and $\hat{\tau}^2 \approx 0.030$ in the rest. Four of the five methods supply similar mean estimates of $\bar{\tau}^2$. The final method indicates no idiosyncratic variation in 90% of the regressions, which implies that the structural model of $\ln p$ is exactly as described by the trend component and that there are no omitted variables. Since that is highly unlikely, especially in view of the evidence from the other methods, I infer that the B estimates of τ^2 are inaccurate and that the other methods supply equally accurate estimates of τ^2.

When spatial correlation decays exponentially the partial sill σ^2 captures the random unexplained variation in price between two properties some distance $\boldsymbol{h}$ apart. Consider a property with mean attributes and sold at mean price US\$77,915.[11] According to the B method another identical property might randomly differ in value by $1km$ away by 3.3%. The corresponding numbers for the C, M, R and H methods are 100%, 0%, 27.4% and 1.2%. The C and R methods imply extremely high levels of instability in the property market. The high value of $\bar{\sigma}^2_C$ implies that one property with mean attributes $1km$ away from the mean priced property might randomly cost virtually nothing while another property with mean attributes $1km$ away would cost twice as much. Such high unexplained variability in the property market is unlikely. I infer that the C and R methods supply biased estimators of σ^2 and that the other methods do not.

Overall the B method supplied a biased mean estimate of one one $\boldsymbol{\alpha}$ parameter and the C and R procedures provide biased mean estimates for two $\boldsymbol{\alpha}$ parameters. Also

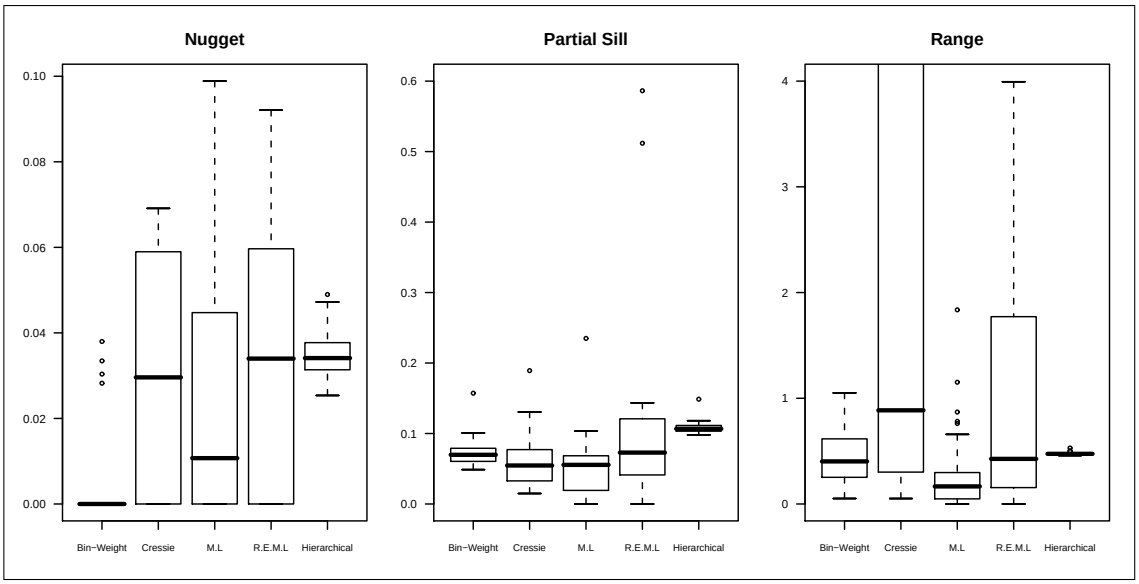
[11] The exchange rate used was US\$1 = 2179 Colombian pesos, valid on 06/30/2009

the bias in the mean R estimates is less than in the mean C estimates. I infer that if accuracy of the $\boldsymbol{\alpha}$ estimators is desirable then the preference ordering over methods is $M \sim H \succ B \succ R \succ C$.

Comparison of $sd(\alpha)$ and $r(\alpha)$ imply big differences in the precision of $\boldsymbol{\alpha}$ estimates across the five methods. The H method supplies by far the most precise estimates. Across methods the relative differences in $sd(\hat{\tau}^2)$ are as follows: $sd(\hat{\tau}^2_H) \approx sd(\hat{\tau}^2_B)/2 \approx sd(\hat{\tau}^2_C)/5 \approx sd(\hat{\tau}^2_M)/5 \approx sd(\hat{\tau}^2_R)/5$. The relative differences in $sd(\hat{\sigma}^2)$ are $sd(\hat{\sigma}^2_H) \approx sd(\hat{\sigma}^2_B)/2 \approx sd(\hat{\sigma}^2_M)/5 \approx sd(\hat{\sigma}^2_R)/90 \approx sd(\hat{\sigma}^2_C)/2550$. Finally, the relative differences in $sd(\hat{\phi})$ are $sd(\hat{\phi}_H) \approx sd(\hat{\phi}_M)/30 \approx sd(\hat{\phi}_B)/250 \approx sd(\hat{\phi}_R)/4450 \approx sd(\hat{\phi}_C)/330,000$. Comparison of the relative differences in ranges across methods of estimation yields similar results. The substantial differences in $sd(\alpha)$ and $r(\alpha)$ suggest that $\bar{\boldsymbol{\alpha}}_H$ is the most precise estimator of $\boldsymbol{\alpha}$ and that $\bar{\boldsymbol{\alpha}}_C$ and $\bar{\boldsymbol{\alpha}}_R$ are the least and second least precise estimators of $\boldsymbol{\alpha}$. With respect to precision in estimation of $\boldsymbol{\alpha}$ the five estimation methods may be ranked $H \succ B \succ M \succ R \succ C$.

The huge disparity in the precision of $\hat{\boldsymbol{\alpha}}$s across procedures and the superior precision of the H procedure are made clear through the box-and-whisker plots in Figure 1.4. The rectangular boxes mark the interquartile range, the thick horizontal line in the boxes mark the median and the whiskers extend out to the most extreme estimate that is no more than 1.5 times the interquartile range from the box. The dots beyond the whiskers mark outliers. The H estimates of all error parameters have much smaller ranges than the others. The ranges for the C and R estimates are especially high. The true difference in ranges is not apparent in Figure 1.4 because the many of the highest estimates are not shown. Consider the median estimates of $\boldsymbol{\alpha}$ as marked by the thick horizontal lines

Fig. 1.4. Box-and-Whisker Plots of Error Parameter Estimates



in Figure 1.4. In support of the assertion in Schabenberger and Gotway (2005, pg. 261) the nugget chart shows a downward bias in the M method median estimates of τ^2. Also, note that the median value of $\hat{\tau}^2_B$ is strongly biased downward just like $\bar{\tau}^2_B$ was.

Finally, using the special case $\hat{\alpha} = \bar{\alpha}$ the mean-squared errors of the error coefficients are identical to their variance. Since $MSE(\bar{\alpha})$ is a monotonic transformation of $sd(\hat{\alpha})$ inferences about the precision of $\hat{\alpha}$ are applicable to the mean quality of $\bar{\alpha}$. By induction, the H method supplies the highest quality estimators of α and the C method supplied the poorest quality estimators. If the quality of $\boldsymbol{\alpha}$ estimators is important then preferences over the five methods of estimation will be defined as $H \succ B \succ M \succ R \succ C$.

1.4.4 Price Predictions

Unlike with the model coefficients, the exact accuracy of price predictions is known because the sales prices of the properties are known. Prediction bias is $b(p, \hat{p}) = E(\hat{p}) - p$, where $\hat{p}$ is the estimator of house price p. The plug-in estimator of $E(\hat{p})$ is the mean prediction $\bar{p} = \sum_{k=1}^{38} \hat{p}_k/38$, where $\hat{p}_k$ is the kth prediction estimate of p. Hence estimated prediction bias is $\hat{b}(p, \hat{p}) = \bar{p} - p$. The smaller the estimated bias, the more accurate the prediction.

Mean bias $\bar{p}$ is reported as the un-bracketed numbers in Table 1.5. Predictions made using the C method had the lowest mean bias at six locations while predictions made using the B and M methods has the lowest mean bias at three locations. The H and R methods generated predictions with the lowest mean bias at two and one location(s) respectively.

Table 1.5: Prediction Biases and Standard Errors

Location	B	C	M	R	H
A	-0.013^{*}	0.000^{*}	0.014^{*}	0.029^{*}	0.032^{*}
	(0.126)	(0.157)	(0.101)	(0.109)	(0.116)
B	0.034^{*}	0.045^{*}	0.029^{*}	0.037^{*}	$0.053^{\dagger}$
	(0.195)	(0.202)	(0.128)	(0.138)	(0.140)
C	0.349	0.364	0.359	0.363	0.361
	(0.153)	(0.175)	(0.117)	(0.119)	(0.132)
D	-0.073	$-0.066^{\dagger}$	-0.075	-0.068	$-0.054^{\dagger}$
	(0.147)	(0.162)	(0.110)	(0.115)	(0.134)
E	0.010^{*}	0.002^{*}	0.012^{*}	0.005^{*}	0.015^{*}
	(0.160)	(0.174)	(0.112)	(0.114)	(0.146)
F	0.391	0.362	0.403	0.382	0.367
	(0.153)	(0.186)	(0.150)	(0.146)	(0.152)
G	-0.322	-0.342	-0.354	-0.375	0.376
	(0.190)	(0.205)	(0.147)	(0.160)	(0.181)
H	-0.111	-0.077	-0.147	-0.134	-0.099
	(0.137)	(0.165)	(0.090)	(0.102)	(0.110)
I	-0.495	-0.473	-0.533	-0.529	-0.519
	(0.141)	(0.164)	(0.091)	(0.100)	(0.120)

Location	B	C	M	R	H
J	0.120	0.135	0.087	0.099	0.140
	(0.167)	(0.167)	(0.117)	(0.117)	(0.169)
K	0.158	0.157	0.141	0.146	0.148
	(0.178)	(0.188)	(0.151)	(0.158)	(0.167)
L	0.599	0.607	0.620	0.615	0.602
	(0.146)	(0.182)	(0.120)	(0.126)	(0.154)
M	$-0.063^{\dagger}$	-0.053^{*}	$-0.047^{\dagger}$	-0.039^{*}	-0.030^{*}
	(0.159)	(0.217)	(0.131)	(0.184)	(0.210)
N	0.143	0.128	0.127	0.110	0.115
	(0.191)	(0.206)	(0.169)	(0.174)	(0.175)
O	-0.244	-0.204	-0.263	-0.247	-0.249
	(0.096)	(0.163)	(0.088)	(0.098)	(0.120)

† and * indicate that the bias is nonsignificant at the 5% and 1% levels. All other biases are significant at the 1% level.

At the other end of the spectrum, $\hat{b}_M$ was greatest at six locations, while the $\hat{b}_H$, $\hat{b}_B$ and $\hat{b}_C$ were greatest at five, three and one location(s) respectively. By thinking of absolute bias as a score the Wilcoxon Signed Rank test can be applied to determine if the differences in bias are significant. The Wilcoxon tests show that mean bias in price predictions is not significantly different across the different methods of estimation. With respect to accuracy in price prediction the preference ordering across the methods of estimation is $B \sim C \sim M \sim R \sim H$.

However there is a spatial pattern to the biases. Comparing Figures 1.1(b) and 1.2, one notes that locations A to I are sited in areas with a high density of sample units and Locations J to O are spatial outliers. Applying the Wilcoxon test to prediction bias in the non-outliers I find that bias when using the C method is significantly less than when using the Maximum Likelihood and Bayesian methods. Bias when using the B method is significantly less than when using the M method. Significance is calculated at the 10% level. Since the C method is significantly more accurate than three other methods, the B method is significantly more accurate than one other method, the R and

H methods are significantly less accurate than one other method and the M method is significantly less accurate than two other methods, one may infer that the methods may be ranked $C \succ B \succ R \sim H \succ M$ when accurate prediction of non-outliers is the criterion. Among the outliers there is no significant difference in prediction accuracy across methods.

Prediction biases at locations I and L are highest across methods. I is not an outlier, which indicates that spatial effects are not to blame for the poor prediction. The data shows that the property at I has an unusually high price and the property at L has an unusually low price. The model specification was unable to capture the price idiosyncrasy at these locations. Instead, the model predicted prices that tended towards the mean, implying a large downward bias at I and a large upward bias at L.

The standard deviation and range of price predictions are used to compare the precision of the different methods. The standard deviation in prediction estimates is $sd(\hat{p}) = \sqrt{\sum_k (\hat{p}_k - \bar{p})^2/38}$, and the range is $r(\hat{p}) = \max(\hat{p}_k) - \min(\hat{p}_k)$ where k indexes the subsamples. The standard deviations of prediction estimates are reported as bracketed numbers in Table 1.5 and ranges are reported as un-bracketed numbers in Table 1.6. Since sd and r results with respect to the relative precision of the different methods is identical they are reported together. $sd(\hat{p}^M)$ and $r(\hat{p}^M)$ are lowest at 14 locations and $sd(\hat{p}^C)$ and $r(\hat{p}^C)$ are highest at 14 locations. $sd(\hat{p}^R)$ and $r(\hat{p}^R)$ are second lowest at 12 locations, $sd(\hat{p}^H)$ and $r(\hat{p}^H)$ are third lowest at 11 locations and $sd(\hat{p}^B)$ and $r(\hat{p}^B)$ are second highest at 11 locations. Wilcoxon tests indicate that sd and r under the M method are significantly less than under all other methods. The tests also indicate that sd and r under the R method are significantly less than under all other methods

except M. The standard deviation and range of prediction estimates under the B and H methods are not significantly different from each other, but are significantly lower than under the C method. From the most precise to the least precise, the methods of estimation may be ordered as follows: $M \succ R \succ B \sim H \succ C$.

Among the non-outliers at locations A to I in Figure 1.2 the relative precision of prediction estimates across methods is similar to relative precision when all locations are considered. The only difference is that the precision under the H method is significantly less than under the B method. The ordering of methods from least to most precise is $M \succ R \succ H \succ B \succ C$. Among the outliers at locations J to O the M method is significantly more precise than the other methods, R is significantly more precise than H and C, and B and H are significantly more precise than C. The precision ordering from least to most is $M \succ R \succ B \succ H \succ C$. The H method is significantly more precise than the B method at prediction of non-outliers, but the opposite holds in the case of outliers.

Prediction quality is gauged through the Estimated Mean Squared Error $EMSE(\hat{p}) = V(\hat{p}) + \hat{b}(p, \hat{p})^2$. The lower the $EMSE$ the better the estimator. The $EMSE$ results are reported as the bracketed numbers in Table 1.6.

Table 1.6: Prediction Range and Expected MSE

Location	B	C	M	R	H
A	0.496	0.614	0.396	0.428	0.455
	(0.016)	(0.025)	(0.010)	(0.013)	(0.014)
B	0.765	0.790	0.502	0.539	0.548
	(0.039)	(0.043)	(0.017)	(0.020)	(0.022)
C	0.601	0.688	0.460	0.467	0.518
	(0.145)	(0.163)	(0.143)	(0.146)	(0.148)
D	0.576	0.634	0.431	0.452	0.524
	(0.027)	(0.031)	(0.018)	(0.018)	(0.021)

Location	B	C	M	R	H
E	0.629	0.683	0.438	0.445	0.571
	(0.026)	(0.030)	(0.013)	(0.013)	(0.021)
F	0.599	0.730	0.589	0.571	0.597
	(0.176)	(0.166)	(0.185)	(0.167)	(0.158)
G	0.745	0.804	0.577	0.629	0.709
	(0.140)	(0.159)	(0.147)	(0.166)	(0.175)
H	0.539	0.646	0.352	0.400	0.430
	(0.031)	(0.033)	(0.030)	(0.028)	(0.022)
I	0.553	0.643	0.356	0.394	0.470
	(0.265)	(0.251)	(0.292)	(0.290)	(0.284)
J	0.654	0.656	0.457	0.459	0.662
	(0.042)	(0.046)	(0.021)	(0.023)	(0.048)
K	0.697	0.738	0.593	0.621	0.654
	(0.057)	(0.060)	(0.043)	(0.046)	(0.050)
L	0.571	0.715	0.469	0.472	0.603
	(0.380)	(0.402)	(0.399)	(0.394)	(0.387)
M	0.624	0.851	0.515	0.720	0.823
	(0.029)	(0.050)	(0.019)	(0.035)	(0.045)
N	0.750	0.807	0.664	0.684	0.685
	(0.057)	(0.059)	(0.045)	(0.043)	(0.044)
O	0.378	0.640	0.344	0.386	0.471
	(0.069)	(0.069)	(0.077)	(0.072)	(0.076)

The M method predictions have the lowest $EMSE$ at eight of 15 locations, the B, C and H methods have the lowest $EMSE$ at two locations each and the R method has the lowest $EMSE$ at one location. The R method has the lowest quality predictions at ten locations, the M method performs worst at three locations and the H method at two. The Wilcoxon tests indicate that the quality of the predictions using the C method are significantly worse than the quality of the predictions from the B, M and R methods at the 5% level. For the H method the significance level is 6%. There is no significant difference in the quality of the predictions from the other methods. The ordering of the methods with respect to prediction quality is $B \sim H \sim M \sim R \succ C$. However, when only considering the non-outliers at locations A to I there is no significant difference in prediction quality across the five methods: $B \sim H \sim M \sim R \sim C$. Prediction quality among the outliers is significantly worse when using the C method, while there

is no difference between the other methods. I infer that when prediction is required at a location that does not lie on the edges of the geographical window then any of the five tested methods of estimation will suffice. When the prediction location is an outlier, on the other hand, then the C method should be avoided.

1.5 Conclusion

The orderings across the different methods of estimation, when considering the accuracy, precision or quality of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ or $\boldsymbol{p}$ estimates, are summarized in Table 1.7. In the table the best performer is assigned a score of one and the worst a score of five. When two or more methods perform equivalently then they share the mean rank. For example, when considering accuracy in the estimation of $\boldsymbol{\alpha}$, note that the H and M methods are better than the others, but there is no significant difference in their performance. Hence both methods are assigned a score of 1.5. The results indicate that

Table 1.7. Relative Performance of Methods of Estimation

		B	C	M	R	H
Accuracy	$\boldsymbol{\beta}$	3.0	3.0	3.0	3.0	3.0
	$\boldsymbol{\alpha}$	3.0	5.0	1.5	4.0	1.5
	$\boldsymbol{p}$	3.0	3.0	3.0	3.0	3.0
Precision	$\boldsymbol{\beta}$	3.0	5.0	1.0	3.0	3.0
	$\boldsymbol{\alpha}$	2.0	5.0	3.0	4.0	1.0
	$\boldsymbol{p}$	3.5	5.0	1.0	2.0	3.5
Quality	$\boldsymbol{\beta}$	3.0	5.0	1.0	3.0	3.0
	$\boldsymbol{\alpha}$	2.0	5.0	3.0	4.0	1.0
	$\boldsymbol{p}$	2.5	5.0	2.5	2.5	2.5

the suitable choice of estimation method is contingent on the research question. The Hierarchical Bayesian method is most suited to an analysis of the error structure of a

spatial hedonic model since it supplies estimators of $\boldsymbol{\alpha}$ that are the most accurate and precise. The Classical Maximum Likelihood method is most suited to an analysis of the trend structure because its estimates of $\boldsymbol{\beta}$ are as accurate as the other methods and the most precise. For prediction on the other hand, all methods are equally suitable, with the exception of Cressie's Least Squares based method. Contrary to the suggestion in Cressie (1993), the $IWLS$ approach with the Cressie Loss function (1.7) did not supply estimates of $\boldsymbol{p}_0$ that were more accurate than the other methods.

If overall accuracy across $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{p}$ is of maximum importance then the M and H methods are most suitable. The mean accuracy score of these methods is 2.5, compared to scores of 3.0, 3.3 and 3.7 for the B, R and C method. The accuracy of the M method belies the expectation that the Maximum Likelihood based approaches will perform poorly with respect to accuracy (Cressie 1993; Schabenberger and Gotway 2005). If overall precision is most important then the M method should be used. Its mean precision score is 1.67, compared to scores of 2.5, 2.8, 3.0 and 5.0 for the H, B, R and C methods. When considering overall quality, the H and M methods have a score of 2.2, the B method has a score of 2.5 and the R and C methods have scores of 3.2 and 5.0 respectively. The Hierarchical Bayesian and Classical Maximum Likelihood methods supply estimates with the highest overall quality when applied to spatial hedonic models. The IWLS method with the Bin-Weighted Loss function (1.6) also also supplied relatively high quality estimates. The Restricted Likelihood method also performs relatively poorly and Cressie's method supplies estimates of the lowest quality and should be avoided.

This paper is only a preliminary stab at comparing different estimation frameworks in a spatial hedonic context. Yet, given the paucity of empirical analyses of

different frameworks in a spatial context, I hope that this research will contribute in some small way to the broader discourse on methodology in spatial statistics. There are many ways in which this work can be extended. Most obviously, the results can be made more robust by increasing the number of n-dimensional realizations from the data set. Further testing of the different procedures is always possible through the design of more metrics. Other estimation approaches can also be tested.

Chapter 2

Water Quality Trading when Nonpoint Pollution Loads are Stochastic

Beginning with the seminal works of Crocker (1966), Dales (1968) and Montgomery (1972), a large literature has developed on the use of emissions trading to achieve environmental targets. The economic case for emissions trading is that it can achieve environmental targets at lower social cost than traditional design and performance standards and emissions taxes (Tietenberg 1990). Real world success stories for air emissions trading have spurred interest in expanding the scope of emissions trading to water-based pollutants (Shortle and Horan 2008).

Water quality trading was the focus of John Dales' (1968) seminal book recommending markets for environmental management, but has not until recently been a focus of applications of the tool. Interest is now high. Dozens of water quality trading initiatives are underway for nutrients, sediments and other pollutants in the US and other countries (Shortle and Horan 2008). The economic appeal of water quality trading is, as is the appeal of emissions trading generally, that it offers a means for achieving a cost-effective allocation of allowable emissions among alternative pollution sources without environmental regulators knowing the abatement costs of individual agents. This promise, if markets can in fact be designed to realize it, is compelling for water pollution control in the US (Shortle and Horan 2008). Water pollution there has largely been regulated through non-tradable, technology-based effluent permits applied to point sources

of water pollution, while nonpoint sources have been largely unregulated (Ribaudo et al. 2005). The economic consequence is that the pollution control that is achieved, because it does not consider the relative costs of alternative point and nonpoint sources, is overly costly (Davies and Mazurek 1998; Ribaudo et al. 2005; USEPA 2001). Studies show that water quality trading programs could result in annual savings of US$ 1 billion when compared to the traditional command-and-control type technology standards (USEPA 2001).

Yet, while the promise of water pollution trading may be compelling, significant challenges confront the design of markets that can realize it. Some key assumptions underpinning the standard emissions trading model are that the polluter has deterministic control over the loadings, and that they are cheaply and accurately measurable (Malik et al. 1993). In the context of water quality trading, these assumptions are valid for point sources but not for nonpoint sources (Shortle 1987). Nonpoint water pollution is largely the result of runoff from agricultural land and other uses. Farmers and other types of land owners cannot control the quality and quantity of runoff with precision; their abatement activities are best characterized as influencing the probabilistic distribution of emissions from their land (Segerson 1988; Horan and Shortle 2001). Further, the dispersed nature of nonpoint pollution makes cheap and accurate metering of emissions impossible. Observational uncertainty compounds the inherent stochasticity of the process. These features of nonpoint pollution raise two significant and related issues for market design. The first is what to define as the nonpoint commodity. The second is how to design trading rules to manage nonpoint trading risk.

The optimal basis for nonpoint instruments, whether for trading or other incentive or regulatory mechanisms, has been much discussed in the economic literature on nonpoint pollution (Shortle and Horan 2001). The choices that have drawn attention for incentive design generally are observable inputs or practices that affect nonpoint pollution (e.g. fertilizer applications to farm land, farming practices that affect sediment or nutrient runoff), emissions estimates constructed from observations on input and practices (e.g. estimates of pollution delivery from farm fields using observations of inputs and practices as data), and realized ambient pollution loads. Tax / subsidy schemes based on realized ambient loads have been of particular interest to economists as a means for nonpoint pollution control (e.g., Segerson 1988; Cabe and Herriges 1992; Horan et al. 2002), but do not offer a plausible basis for nonpoint pollution trading because trading involves an exchange of individual responsibilities, while ambient loads measure group performance. Thus, policy interest has largely focused on trading inputs and practices, or estimated pollution loads.

The approach in US water quality trading programs is to define the tradable nonpoint commodity in terms of estimated reductions in emissions or delivered loads (Shortle and Horan 2008). An important aspect of this approach should be noted in comparison to conventional trading of actual metered emissions. This is the enormous uncertainty about the actual water quality outcomes of individual trades based on modeled emissions. This uncertainty exists because the prediction errors for water quality models are known to be quite large (National Research Council 2001; Reckhow 1994, 2003). This uncertainty is addressed in US program design by the application of an "uncertainty" trading ratio to nonpoint trades. The ratio is typically defined in terms

of the reduction in nonpoint emissions required to offset a unit of point source emissions. Designing uncertainty trading ratios to address nonpoint risk has been the focus of the limited economic research on point-nonpoint trading (e.g., Hennessy and Feng 2008; Rabotyagov et al. 2006; Malik et al. 1993; Shortle 1990). Three essential insights have resulted (Shortle and Horan 2008). One is that trading ratios are optimally differentiated across sources to address differences in relative risk. A second insight is that optimal trading ratios defining the exchange rate between nonpoint discharges and point source emissions may be less than one, reflecting the fact that nonpoint sources are often a greater source of risk than point sources. A third insight is that optimal trading ratios depend on the choices of other market design parameters, including the allocation of initial allowances and the cap placed on the aggregate supply of permits, and thus optimality required selecting these parameters simultaneously. These findings are contrary to standard practice in which point-nonpoint trading ratios are almost always uniform across nonpoint sources or categories of nonpoint sources, often (substantially) more than one, and specified independently of other parameters (Horan 2001; Shortle and Horan 2008).

The economic literature thus suggests flaws in the design of actual trading programs that will affect their capacity to minimize costs and optimally manage water quality risk. This paper will argue that those flaws aside, the use of modeled mean emissions as an approximation of nonpoint source pollution and the use of trading ratios to address nonpoint risk is suboptimal, and will offer an alternative market design that offers greater promise for efficiency in water quality management through trading.

Emissions markets are conventionally designed to meet an exogenously specified cap on a linear aggregation of polluting emissions. As noted above, water quality planners are concerned with risk. When emissions are stochastic, an alternative to the conventional design is to place a limit on the probability that the linear aggregation exceeds the exogenously specified cap. Probabilistic environmental targets of this type are labeled "safety-first" and are consistent with the Total Maximum Daily Load ($TMDL$) approach to water quality management mandated by the Clean Water Act (Beavis and Walker 1983; Lichtenberg and Zilberman 1988; Horan et al. 2002; Qiu et al. 2001; National Research Council 2001).

In this paper I develop the concept of a safety-first environmental target in the context of water quality management and derive necessary conditions for the attainment of this target at least cost. I then propose a market design and show that at equilibrium the market outcome meets the safety-first target at least cost. The primary innovation in market rules, when compared to conventional environmental markets for nonstochastic pollutants in which polluters trade emissions allowances, or US water quality markets in which nonpoint sources trade estimated emissions-based credits, is in the definition of the nonpoint source credits. The nonpoint source credits traded in this market have two attributes: the mean and variance of abatement. I also study the performance of existing trading ratio-based water quality trading markets vis-à-vis their attainment of the safety-first environmental target.

On the surface there are similarities between the market with the multi-attribute nonpoint credits and the CAPM market in the finance literature since both are designed to control variance and risk. A closer inspection reveals differences in the details. Unlike

with any environmental market there is no broad policy intent behind the CAPM market. The existence of the market is taken as given and the focus of the CAPM analysis is mean-variance efficiency as a solution concept. An environmental market on the other hand is brought into existence by policy dictate and the solution concept centers around the attainment of the environmental target. By making the safety-first environmental target the starting point of my analysis, model development differs from CAPM model development. Additionally some of the assumptions of the CAPM model are not made in my analysis. For example I do not assume that nonpoint emissions are normally distributed and there is no space for interest rate differentials in the model. There are similarities in results across the two models, arising from a congruence of interests related to risk control.

In the next section I formally derive the safety-first environmental target and then obtain conditions on polluter behavior required for its least cost attainment. The intent is to generate a clear picture of the optimal scenario. In the section after that, the market with the multi-attribute nonpoint abatement credits, referred to as the MANA market in the subsequent discussion, is described. Market rules are defined and necessary conditions for equilibrium discussed. Next I examine restrictions that must be imposed to ensure that the market equilibrium is safety-first compliant at least cost. The next section focuses on the trading-ratio-based market, which is conceptually identical to water quality trading markets in current use. This market is referred to as the TR market for short. The conditions required for the TR market equilibrium to be safety-first compliant at least cost are derived. In the final section, the MANA market and the TR market are compared, and the results put in a broader policy context.

2.1 The *Safety-First* Environmental Target

The Environmental Agency's problem is to design a market that will achieve its safety-first objective at least cost. Following Shortle and Horan (2008) the Agency has three integrated tasks when pursuing this objective. The first is to *define the commodities that will be traded in the market.* A key element of the definition is the specification of the observable indicators of environmental performance (e.g. pounds of total nitrogen annually discharged from a farm or sewage treatment facility in the Susquehanna River Basin) to which rights pertain. The indicators must be observable so that trading is enforceable, and under the control of the polluter if the polluter is held responsible for non-compliance. The second task is to *define specific rules for trading the commodities between alternate sources.* The defined commodities may not be homogenous across sources. For example, discharges of nitrogen at one point in a watershed may have different water quality impacts than discharges at another. In current markets trading rules are generally implemented in the form of trading ratios that define the rates at which different commodities can be exchanged between different sources. From the perspective of water quality agencies, trading ratios are intended to assure that water quality outcomes resulting from commodity trades are at least equal to those that would occur without the trade (USEPA, 2007). The third task is to *cap the aggregate supply of the commodities such that feasible market allocations of polluting emissions. given the trading rules, do not violate the environmental goal(s).*

I motivate the model by formally defining the Environmental Agency's problem. Following the EPA's watershed based approach to controlling water quality (see USEPA

2003), I situate the problem at the same geographic scale. Consider a representative watershed. The Environmental Agency is tasked with designing a policy that targets loadings from polluters such that all water bodies in the watershed meet their designated use. This requires that ambient aggregate loadings in the watershed be controlled, which makes ambient loadings a policy variable of interest. Let L be ambient loadings and let $\bar{L}$ be the watershed level loadings target. Following the safety-first literature (e.g., Lichtenberg and Zilberman 1988; Qiu et al. 2001) the environmental target is expressed probabilistically. The Agency is required to ensure that the probability that ambient loadings exceed the ambient target is less than α or $\Pr(L \geq \bar{L}) \leq \alpha$. I assume that α is a fixed coefficient set externally and not controlled by the Environmental Agency. If $\alpha = 0.05$ then the safety-first environmental target is set such that the ambient loadings target is exceeded in 5% of all possible scenarios.

Let the market consist of I nonpoint sources and J point sources, indexed by i and j respectively. Let r_i be i's runoff or loadings and let e_j be j's emission loadings into the watershed in a given trading period. Summing over all i and j, aggregate emissions loadings are $L = \sum_i r_i + \sum_j e_j$ and the probabilistic statement of the safety-first environmental target is $\Pr(\sum_i r_i + \sum_j e_j \geq \bar{L}) \leq \alpha$. If the joint distribution of emissions is known, and if the distribution is such that a "deterministic equivalent" exists, then the probability statement above can be expressed exactly in terms of means, variances, covariances and other moments of the distribution of the emissions (Beavis and Walker 1983; Wets 1983; Kampas and White 2003). But given the nonmeasurability of nonpoint emissions and the prediction errors in current water quality models, knowledge

of the emissions distribution will not be known with certainty and that the conditions for a deterministic equivalent will not be satisfied.

A common nonparametric approximation of probability statements when the distribution of the random variable is unknown (or a deterministic equivalent does not exist) is based on Chebychev's inequality (Wets 1983). By Chebychev's inequality the statement $\Pr\left(x \geq E(x) + \sqrt{V(x)/\alpha}\right) \leq \alpha$ is always true, where x is a random variable, $E(x)$ and $V(x)$ are its mean and variance, and $\alpha \in [0, 1]$ is a probability. Let μ_i be the mean value of nonpoint source i's runoff r_i and let σ_i^2 be its variance. Taken together μ_i and σ_i^2 provide a second order approximation of the stochasticity of i's emissions, and thus characterize the risk that i's pollution activities pose vis-à-vis the attainment of the safety-first target. If the true distribution of r_i were normal then μ_i and σ_i^2 completely define the probability distribution of r_i. In general, the true distribution is not normal and approximation errors will exist. It is reasonable to assume that the stochasticity of r_i is the cumulative effect of many non-normal random processes. Since the Central Limit Theorem indicates that the sum effect of many non-normal random processes is approximately normal (Banerjee et al. 2004), one may assume the same for r_i. In which case, one may expect the approximation error to be small.

For the purposes of this analysis I assume that nonpoint source emissions are independent, which means that $\text{cov}(r_i, r_l) = 0$, where $i, l \in I$. This assumption is unlikely to hold for contiguous polluters who experience similar weather and topography. It is more realistic in the case of non-contiguous polluters facing different micro-weather patterns. The assumption is used because it simplifies the ensuing analysis. To apply Chebychev's inequality to the safety-first context, set $x = \sum_i r_i + \sum_j e_j$, which implies

that $E(x) = \sum_i \mu_i + \sum_j e_j$ and $V(x) = \sum_i \sigma_i^2$ and

$$\Pr\left(\sum_i r_i + \sum_j e_j \geq \sum_i \mu_i + \sum_j e_j + \sqrt{\frac{\sum_i \sigma_i^2}{\alpha}}\right) \leq \alpha \tag{2.1}$$

Comparing (2.1) to the probabilistic statement of the safety-first environmental target, $\Pr(\sum_i r_i + \sum_j e_j \geq \bar{L}) \leq \alpha$, it is apparent that the target is satisfied when

$$\sum_i \mu_i + \sum_j e_j + \sqrt{\frac{\sum_i \sigma_i^2}{\alpha}} \leq \bar{L} \tag{2.2}$$

This formulation of the safety-first environmental target is similar to that in Beavis and Walker (1983). Equation (2.2) indicates satisfying the safety-first constraint requires that nonpoint sources control the mean and variance of their emissions runoff and that point sources control their measured emissions. Reductions in μ_i, σ_i^2 and e_j weaken the constraint and are thus desirable actions from the Environmental Agency's point of view. Indeed, μ_i, σ_i^2 and e_j may be interpreted as substitutes in the satisfaction of the safety-first environmental target. When (2.2) binds then μ_i and e_j should be treated as perfect substitutes. The marginal rate of substitution between these variables and σ_i^2 is $2\sqrt{\alpha \sum_i \sigma_i^2}$, which is large when there are many nonpoint sources in the watershed. These results are obtained by treating (2.2) as an equality and differentiating both sides with respect to all variables. Since $2\sqrt{\alpha \sum_i \sigma_i^2}$ units of σ_i^2 are equivalent to one unit of μ_i or e_j at the margin, one may infer that control of the latter is suitable for rapid movement towards the safety-first outcome. Once in the vicinity of the safety-first outcome however, control of σ_i^2 allows minute calibration towards the precise satisfaction of (2.2).

2.1.1 The Least Cost Safety-First Solution

Assume that the controlling μ_i and σ_i^2 affects nonpoint source i's profits and that controlling e_j affect point source j's profits. Let $\pi_i(\mu_i, \sigma_i^2)$ and $\pi_j(e_j)$ be the restricted profit functions for representative nonpoint source i and point source j respectively. Assume that π_i and π_j are increasing and concave in all arguments for all polluters. Restricted profit functions define the maximum profits associated with any emissions level, given that the input mix and output levels are chosen optimally (Graff Zivin and Small 2003). I derive the restricted profit functions $\pi_i(\mu_i, \sigma_i^2)$ and $\pi_j(e_j)$ from the standard profit function, where profits are a function of inputs and prices, in Appendix A.1. Under the restricted profit formulation note that the nonpoint source profit function is not stochastic. Rather, deterministic nonpoint source profits and stochastic emissions are associated with the deterministic vector of inputs. The stochastic emissions are approximated through the known and controllable moments μ_i and σ_i^2. Then, through the restricted profit function, a relationship is generated between deterministic profits and the known moments.

The aggregate restricted profit function for all polluters in the watershed is $\sum_i \pi_i(\mu_i, \sigma_i^2) + \sum_j \pi_j(e_j)$. I assume that the Environmental Agency wants polluters to choose μ_i, σ_i^2 and e_j for all $i, j \in I \times J$ such that aggregate profits are maximized subject to the safety-first environmental target (2.2). This solution implies the least-cost attainment of the safety-first environmental target and is optimal from the Agency's point of view. Since, as stated in Appendix A.1, positive production levels are accompanied by positive emissions levels, it must be that all choice variables, $\{\mu_i, \sigma_i^2, e_j\}_{i,j}$,

are strictly positive. Given this qualification, the conditions that are necessary for the Agency's optimum are

$$\frac{\partial \pi_j}{\partial e_j} = \frac{\partial \pi_i}{\partial \mu_i} = 2\frac{\partial \pi_i}{\partial \sigma_i^2}\sqrt{\alpha \sum_i \sigma_i^2} = \lambda \quad \forall\, i, j \tag{2.3}$$

where λ is the Lagrange multiplier and shadow price associated with the safety-first constraint in (2.2). The conditions in (2.3) were obtained through manipulation of the Kuhn-Tucker conditions. If a vector $\{\mu_i, \sigma_i^2, e_j\}_{i,j}$ that satisfies (2.2) forms a convex set then (2.3) is also sufficient conditions for the least-cost attainment of the safety-first environmental target. This is true because of the concavity of π_i and π_j in all arguments.

The first equality in (2.3) indicates a unitary marginal rate of substitution between μ_i and e_j, $MRS_{\mu_i, e_j} = 1$, at the least cost safety-first solution. This implies that μ_i and e_j are treated as perfect substitutes with a one-to-one correspondence. The second equality establishes the optimal trade-off between μ_i and σ_i^2. The marginal rate of technical substitution between μ_i and σ_i^2 is $MRTS_{\mu_i, \sigma_i^2} = 2\sqrt{\alpha \sum_i \sigma_i^2}$, where $\sum_i \sigma_i^2$ is the aggregate variance of nonpoint source abatement. $\sqrt{\alpha \sum_i \sigma_i^2}$ may be interpreted as the aggregate risk in the market, after adjusting for society's appetite for environmental risk, as defined by α. The $MRTS$ result implies that at the optimum solution a unit increase in μ_i is accompanied by a decrease in σ_i^2 by $2\sqrt{\alpha \sum_i \sigma_i^2}$ units and vice versa. Consider an example where $\sum_i \sigma_i^2 = 500$ and $\alpha = 5\%$ at the least cost solution. From (2.3) remaining at least cost requires $MRTS_{\mu_i, \sigma_i^2} = 10$. A unit increase in μ_i must be accompanied by a ten unit decrease in σ_i^2 to keep aggregate profits at their minimum. Vis-à-vis the maintenance of Agency optimality, σ_i^2 is a more sensitive control variable

than μ_i. The results on the use of $\{\mu_i, \sigma_i^2, e_j\}_{i,j}$ to satisfy (2.2) transfer unchanged to the least cost solution: μ_i and e_j can be adjusted to move towards the least cost safety-first solution and then σ_i^2 can be used for fine-tuning in the neighborhood of the optimum. The final equality in (2.3) indicates that marginal profits with respect to μ_i and e_j and risk-adjusted marginal profits with respect to σ_i^2 must be equated to the shadow price of the safety-first constraint, λ.

2.2 The MANA Market

Following Shortle and Horan (2008), the safety-first environmental constraint (2.2) and the least cost solution satisfying this constraint, as described in (2.3), suggest that explicit control of measured point source emissions and the modeled mean and variance of nonpoint source emissions is necessary for the attainment of the environmental target. One way to achieve this aim is to define the traded commodities in terms of these variables and then impose market rules that facilitate the desired outcomes.

2.2.1 Definitions and Rules

I define an allowance as a right to pollute. It identifies the maximum emissions that an individual polluter may discharge into the watershed. It is allocated by the Environmental Agency and known to the polluter at the start of the trading period. Let $\hat{r}_i$ and $\hat{e}_j$ be the allowances given to a representative nonpoint source i and a representative point source j respectively. I assume that the vector of allowances $\{\hat{e}_1, \cdots, \hat{e}_j, \cdots, \hat{e}_J, \hat{r}_1, \cdots, \hat{r}_i, \cdots, \hat{r}_I, \}$ account for distance effects. Distance-related trading ratios or "delivery" ratios, which control for spatial impacts on ambient

loadings, are not explicitly modeled. Implicit in this treatment of spatial effects is that they are known and non-stochastic. The allowances are allocated such that they sum up to the watershed level pollution target $\bar{L}$ or $\sum_j \hat{e}_j + \sum_i \hat{r}_i = \bar{L}$. The allowances are parameters outside the control of the polluters.

2.2.1.1 Credits

The difference between nonpoint source i's allowance and runoff is $a_i = \hat{r}_i - r_i$, which is unknown because of the stochasticity and non-measurability of r_i. The expected difference between allowance and runoff is $E_i = \hat{r}_i - \mu_i$ and the variance of the difference is $V_i = \sigma_i^2$. Let the credits generated by a polluter be defined as the difference between its allowance and its runoff or emissions loadings. Since the true difference a_i is unknown I define the credits generated by nonpoint source i as the two-attribute good $\tilde{a}_i$, where the attributes are the mean E_i and variance V_i of the difference variable a_i. The credit definition, $\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$, contains the information needed to make a second order approximation of the probaiblity distribution of the true difference between i's allowance and runoff. Through the definition, information on the mean and variance of i's runoff is transmitted to the market, which is an important design factor since μ_i and σ_i^2 are critical to the definition and least cost attainment of the safety-first environmental target. The difference between point source j's allowance and emissions is $a_j = \hat{e}_j - e_j$, which is known because e_j is accurately measurable. Hence j's credit supply is simply defined as a_j. Note that credit supply is negative when emissions exceed allowance and is positive otherwise. A polluter is in compliance when its credit supply, after adjustment for trade, is weakly positive. I develop this point in Section 2.2.3.

2.2.1.2 Behavior

Let r_i^0 and e_j^0 be nonpoint source i's and point source j's profit-maximizing emissions prior to market entry. I assume that $r_i^0 \leq \hat{r}_i$ and $e_j^0 > \hat{e}_j$. The representative nonpoint source i faces a lenient allowance that does not prevent it from polluting up to its profit maximizing level under autarky. Point source j's allowance, on the other hand, prevents it from polluting up to its profit maximizing level under autarky. The implication is that nonpoint sources have no incentive to buy credits from other polluters. Point sources have incentive to buy credits from nonpoint sources and other point sources if buying credits is less costly than restricting production.

These relationships between allowances and profit-maximizing emissions reflects realities in extant water quality trading markets. Although major contributors to water quality impairment, before the advent of markets nonpoint sources were outside the purview of the pollution control policy regime (USEPA 2002). Since their inception water quality trading markets were treated as viable instruments to control nonpoint pollution without direct regulation. The idea was to use water quality trading markets as a carrot approach to nonpoint pollution control, whereby nonpoint sources are given financial incentives to voluntarily reduce loadings. Because they do not face legislative requirements to reduce loadings, they have no incentives to buy abatements. Point sources on the other hand do face legally binding loadings allowances. Since buying abatements from nonpoint sources is cheaper than investing in their own abatements, they have incentive to buy from nonpoints. I also allow point sources to trade abatements with each other because such trades are allowed in extant markets.

Since nonpoint source pollution is unobservable, a critical design factor is to predicate E_i and V_i on observable emissions reducting technologies or inputs. For example, consider a nonpoint source i that implements a riparian grass buffer. Nonpoint source i's credit supply $\tilde{a}_i(E_i, V_i)$ is determined by the modeled reduction of runoff as a consequence of the buffer. If the model indicates an expected abatement of 30 units with a variance of 100 i's credit supply has the characteristics $\{30, 100\}$. The variance describes the risk associated with the abatement project. High-risk, high-variance projects increase the probability of environmental targets being violated. As is the practice in current markets, nonpoint sources generate credits by implementing one or more abatement-inducing technologies. Assuming a large set of abatement technologies and the feasibility of marginally adjusting the abatement-generating capability of each technology by varying the input mix, the $\{E_i, V_i\}$ space may be considered continuous.

Consider a nonpoint source that implements a set of technologies and generates the credit supply $\tilde{a}_i(E_i, V_i)$. A point source j may buy a $\gamma_{ij} \in [0, 1]$ share of i's credit supply, where $\gamma_{ij} = 0$ implies that j does not trade with i and $\gamma_{ij} = 1$ implies that j has bought i's entire credit supply. Buying a proportion of i's credit supply is equivalent to investing a γ_{ij} share in i's set of technologies. The point source can minimize the costs of meeting its allowance by optimally adjusting its portfolio of shares in the credit-generating abatement technologies implemented by other polluters. I assume that the market is competitive, which implies that nonpoint sources sell their entire credit supply and $\sum_j \gamma_{ij} = 1 \ \forall \ i$. Point sources can buy and sell shares in their own abatement technologies to other point sources. Let $\gamma_{jk} \in [0, 1]$ be the proportion of point source j's generated credits sold to point source k. $\sum_{k \neq j} \gamma_{jk}$ is the total proportion of credits sold

by j to other point sources and γ_{jj} is the proportion of j's credits that remain unsold. The mutual exclusivity of these events dictates that $\sum_k \gamma_{jk} = \sum_{k \neq j} \gamma_{jk} + \gamma_{jj} \equiv 1$.

Given the continuity of the $\{E_i, V_i\}$ space and of the e_j space, it is not necessary to model credit sales as investments in credit-producing projects. Instead, each transaction may be thought of as predicated upon a separate technology. In other words, instead of assuming that point sources j and k have bought 50% shares in a project supplying 10 credits, we may have assumed that they invested in separate projects supplying 5 credits each. The conceptualization of credit sales as investments was a conscious decision to capture the process through which credits are allocated in the real world. Credits are analogous to offsets whereby the point source contracts with a nonpoint source to implement an abatement technology. Conditional on implementation, the Environmental Agency then associates a number of credits with the transaction, which allows the point source to increase its emissions by the credit amount. Recognizing the possibility that point sources may wish to share risks by investing jointly in projects I allow fractional investment.

2.2.2 The Nonpoint Source Problem

Nonpoint source i's credit supply is $\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$. Prior to market entry i was producing emissions at its unconstrained profit maximizing level. Let its unconstrained profits be π_i^0. When producing credits $\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$ its restricted profit function is $\pi_i(\mu_i, \sigma_i^2)$. Its cost of credit generation is $\pi_i^0 - \pi_i(\mu_i, \sigma_i^2)$. Nonpoint source i also earns revenue $q_i(E_i, V_i)$ from sale of $\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$ in the market. Hence its profit function is $\Pi_i = q_i(E_i(\mu_i), V_i(\sigma_i^2)) - [\pi_i^0 - \pi_i(\mu_i, \sigma_i^2)]$. It is assumed that i wants to maximize

its profit. It faces an unconstrained optimization problem since its allowance is never binding with $r_i^0 \leq \hat{r}_i$. Through the implementation of credit generation technologies it controls μ_i and σ_i^2. The following conditions, obtained by manipulating the Kuhn-Tucker conditions, are necessary for i to maximize profits:

$$-\frac{\partial q_i}{\partial \mu_i} = \frac{\partial \pi_i}{\partial \mu_i} \tag{2.4}$$

$$-\frac{\partial q_i}{\partial \sigma_i^2} = \frac{\partial \pi_i}{\partial \sigma_i^2} \tag{2.5}$$

The equalities in (2.4) and (2.5) indicate that i maximizes profits when marginal revenue equals negative marginal production profits. Since $\partial \pi_i / \partial \mu_i \geq 0$ and $\partial \pi_i / \partial \sigma_i^2 \geq 0$ by assumption, it follows that the profit maximum revenues must be decreasing in μ_i and σ_i^2. Since $E_i = \hat{r}_i - \mu_i$ and $V_i = \sigma_i^2$ (2.4) and (2.5) also indicate that q_i should be increasing in mean credit supply and decreasing in its variance. Nonpoint sources are incentivized to supply credits with a high mean and a low variance.

2.2.3 The Point Source Problem

As defined in Section 2.2.1, the representative point source j faces an emissions allowance $\hat{e}_j$. Under autarky it meets its allowance by choosing emissions level e_j such that $e_j \leq \hat{e_j}$. Or, in terms of the credits it generates , $a_j = \hat{e}_j - e_j$, the allowance is met when it has a positive credit supply, $a_j \geq 0$. Since the allowance $\hat{e}_j$ constrains profit-maximizing behavior point source j will augment its allowance by buying credits from other polluters in the market. Or more accurately, by creating a portfolio of shares in the credit-generating projects of other polluters.

Point source j's purchase of shares from other point sources is $\sum_{k\neq j}\gamma_{kj}a_k$ where k indexes the other point sources. Point source j's purchase of shares from nonpoint sources is $\sum_i \gamma_{ij}\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$. The credits sold by j to other point sources is $\sum_{k\neq j}\gamma_{jk}a_j$ and the unsold credits are $\gamma_{jj}a_j$. Summing over these trades and using the tautology that $\sum_k \gamma_{jk} = \sum_{k\neq j}\gamma_{jk} + \gamma_{jj} \equiv 1$, j's total credit stock is $\sum_k \gamma_{kj}a_k + \sum_i \gamma_{ij}\tilde{a}_i$. Given this credit stock, the true difference between allowances bought by j and emission loadings is $\sum_k \gamma_{kj}a_k + \sum_i \gamma_{ij}a_i$. The mean difference is $\sum_k \gamma_{kj}a_k + \sum_i \gamma_{ij}(\hat{r}_i - \mu_i)$ and the variance of the difference is $\sum_i \gamma_{ij}^2\sigma_i^2$.

Even though j's emissions are deterministic I define its mandatory responsibilities, with respect to its trades in credits, probabilistically. This is because meeting the safety-first environmental target at least cost [see (2.3)] requires probabilistic control of stochastic emissions and defining point source responsibilities probabilistically supplies a convenient way to do so. Point source j's allowance constraint is probabilistic because it offsets its deterministic emissions through the purchase of nonpoint source credits that embody stochastic abatement. This specification – of probabilistic regulation of individual polluters – is inconsistent with current US water quality trading regulations. However, it makes intuitive sense when developing a policy that allows offsets from non-deterministic regulations and watershed level environmental targets are defined probabilistically. Current regulation is deterministic and not based on a probabilistic exposition of the environmental problem.

Let the allowable probability that j exceeds its emissions allowance be β. It is not necessary that β be the same as the allowable probability that the market level environmental target be violated, α. The Environmental Agency (and by proxy society)

may have different preferences over aggregate and individual emissions. Intuitively this makes sense: there is no reason for the Agency to care as much about the likelihood of an individual polluter violating its cap as it does about the aggregate cap being violated. An individual cap violation is not necessarily in conflict with attainment of the safety-first environmental target. Under-abatement by some polluters may be accompanied by equal over-abatement by other polluters. Fundamentally, we need the choice of β to be consistent with the overall goal of meeting the safety-first environmental target at least cost. As is shown subsequently, setting $\beta = \alpha$ is not consistent with this goal.

When offsets from only deterministic sources are allowed, the necessary condition consistent with the point source meeting its allowance is $Emissions - Offsets \leq Allowance$, or in terms of credit stock, $\sum_k \gamma_{kj} a_k \geq 0$. Analogous to the safety-first formulation of the watershed level problem, when offsets from stochastic sources are also allowed, the point source is required to control the probability with which its emissions, after accounting for offsets, violates its allowance. In other words, a safety-first environmental policy requires that $\Pr(Emissions - Offsets \leq Allowance) \geq 1 - \beta$, the mathematical expression of which is $\Pr\left(\sum_k \gamma_{kj} a_k + \sum_i \gamma_{ij} a_i \geq 0\right) \geq 1 - \beta$. Applying Chebychev's inequality, j's allowance constraint expressed in terms of the difference between emissions and allowances is

$$\sum_k \gamma_{kj} a_k + \sum_i \gamma_{ij} E_i - \sqrt{\frac{\sum_i \gamma_{ij}^2 V_i}{\beta}} \geq 0 \tag{2.6}$$

Since $a_j = \hat{e}_j - e_j$, $E_i = \hat{r}_i - \mu_i$ and $V_i = \sigma_i^2$, (2.6) may be rewritten in terms of emissions as $\sum_i \gamma_{ij} \mu_i + \sum_j \gamma_{kj} e_k + \sqrt{\sum_i \gamma_{ij}^2 \sigma_i^2 / \beta} \leq \sum_k \gamma_{kj} \hat{e}_k + \sum_i \gamma_{ij} \hat{r}_i$. Comparing

this expression to the watershed level safety-first environmental constraint (2.2) reveals similarities. Note that $\sum_k \gamma_{kj}\hat{e}_k + \sum_i \gamma_{ij}\hat{r}_i$ is the proportion of the watershed level loadings target $\bar{L}$ that j is responsible for. The point source's constraint in terms of emissions is identical to the safety-first constraint in (2.2) except for the weights supplied by the γ shares. These weights, which measure j's preferences for the risks associated with different credit-producing projects, distort the marginal rate of substitution between emissions and moments from what is required to satisfy (2.2). The impacts of these distortions on the market equilibrium are shortly analyzed.

Let us look at the marginal trade-offs between credits and moments as revealed in (2.6). Just as e_j and μ_i are perfect substitutes in the satisfaction of the watershed level constraint, $\gamma_{kj}a_k$ and $\gamma_{ij}E_i$ are perfect substitutes in the satisfaction of (2.6) where $\gamma_{kj}a_k$ is the quantity of credits bought by j from k and $\gamma_{ij}E_i$ is j's purchased share of the mean abatement offered by i. These quantities are controlled through variation in the corresponding γs. The marginal rates of substitution between these commodities and $\gamma_{ij}^2 V_i$ is $-2\sqrt{\beta \sum_i \gamma_{ij}^2 V_i}$ where $\gamma_{ij}^2 V_i$ is j's purchased share of the variance or risk associated with i's credits. The negative marginal rate of substitution implies that $\gamma_{ij}^2 V_i$ is treated as a bad. Increases in $\gamma_{kj}a_k$ or $\gamma_{ij}E_i$ are accompanied by concurrent increases in $\gamma_{ij}^2 V_i$.

Note that the point source j's allowance constraint (2.6) moves enforcement to an ex-ante perspective. Point sources are no longer held liable for actual emissions. A point source is compliant if it buys enough offsets such that it satisfies (2.6), which in turn ensures that the probability that actual emissions exceed the compliance is β. The point source is not culpable for adverse random events that increase actual emissions. In

any case, using actual emissions as a basis for liability is problematic because the actual nonpoint source emissions are not measurable.

Let us now look at j's cost structure. Let its unconstrained profits be π_j^0. When producing a_j credits its restricted profit function is $\pi_j(e_j)$, rewritten as $\pi_j(e_j)$ by making the substitution $a_j = \hat{e}_j - e_j$. The cost of credit generation is $\pi_i^0 - \pi_j(a_j)$, Let p be the unit price of a point source credit. It does not vary across point sources because the point source credit is homogeneous and the market is competitive. The payment that j makes to other point sources for credit purchase is $p\sum_{k\neq j}\gamma_{kj}a_k$. The payment received from credit sale is $p\sum_{k\neq j}\gamma_{jk}a_j$. Using the tautology that $\sum_k \gamma_{jk} \equiv 1$ net payments to other point sources are $p\big[\sum_k \gamma_{kj}a_k - a_j\big]$. Finally j's total payments to nonpoint sources is $\sum_i \gamma_{ij}q_i(E_i, V_i)$. Summing, j's profit function is $\Pi_j = p[a_j - \sum_k \gamma_{kj}a_k] - [\pi_i^0 - \pi_j(a_j)] - \sum_i \gamma_{ij}q_i(E_i(\mu_i), V_i(\sigma_i^2))$.

I assume that j wants to maximize profits, Π_j. It faces a constrained optimization problem, since its emissions are circumscribed by the relationship in (2.6). It optimizes by controlling its credit generation a_j and by adjusting the proportion of other polluters' credits that it owns $\{\gamma_{kj}, \gamma_{ij}\}_{i,k}$. The following conditions, obtained by manipulating the Kuhn-Tucker conditions and using the definition of a_j, are necessary for j to maximize profits:

$$\frac{\partial \pi_j}{\partial e_j} = p \tag{2.7}$$

$$q_i^* = p\left(E_i(\mu_i) - \frac{\gamma_{ij}V_i(\sigma_i^2)}{\sqrt{\beta\sum_i \gamma_{ij}^2 V_i(\sigma_i^2)}}\right) \quad \forall\, i \in I \tag{2.8}$$

The derivation of the conditions described in (2.7) - (2.8) is provided in Appendix A.2. Both conditions are necessary for profit maximizing point sources who trade with both point and nonpoint sources. If j only traded with other point sources then (2.7) is the only necessary condition of relevance. However, if the policy expectation that nonpoint credit generation is cheaper than point source credit generation is met, then it is sub-optimal for any point source to not trade with nonpoint sources. If j traded with only nonpoint sources then (2.7) - (2.8) are necessary for profit maximization under the caveat that p can be established. Point source credit price p will be established as long as there is at least one trade between point sources. Given that the market is competitive, it is reasonable to expect trades among point sources. In the unlikely scenario where there are no trades between point sources in the watershed, p may be set by reference to prices in other tradable permit markets. Other possible scenarios, such as j not trading at all or only selling credits, are sub-optimal given that (2.6) is a binding constraint and point source abatement is expensive, and merit no further discussion.

The first necessary condition (2.7) indicates that j maximizes profits by choosing its emissions level such that marginal profits are equated with the unit price of point source credits. The second set of optimality conditions (2.8) indicates how $i \in I$ must be paid if j is to maximize profits. Note that the RHS of (2.8) can be decomposed into additively separable functions of E_i and V_i with $q_i^* = pE_i - p\gamma_{ij}V_i/\sqrt{\beta \sum_i \gamma_{ij}^2 V_i}$. This separability suggests that E_i should be treated equivalently to a_j in the market, with both being sold for p. The negativity of the pricing term for variances, $-p\gamma_{ij}V_i/\sqrt{\beta \sum_i \gamma_{ij}^2 V_i}$, suggests that the nonpoint source credit $\tilde{a}(E_i(\mu_i), V_i(\sigma_i^2))$ is penalized for variability at j's profit maximizing solution, but at a decreasing rate. In other words, the penalty for

variability is increasing but concave in V_i. The variance term $\gamma_{ij}V_i/\sqrt{\beta\sum_i \gamma_{ij}^2 V_i}$ may be interpreted as the β-adjusted relative risk of i's credits relative to all other credits. This is because $\gamma_{ij}^2 V_i$ is the variance of the fraction of $\tilde{a}(E_i(\mu_i), V_i(\sigma_i^2))$ bought by j from i and $\sum_i \gamma_{ij}^2 V_i$ is the variance of all credits bought by j.

Let us take a closer look at (2.8) in the context of how the profit maximizing behavior of j affects the profit maximizing behavior of other point sources. Consider some $k \in J$ where $k \neq j$. The profit maximizing necessary conditions for k are also (2.7) and (2.8) with k substituted for j. It follows that if both k and j maximize profits then $q_i^* = p\Big(E_i - \gamma_{ij}V_i/\sqrt{\beta\sum_i \gamma_{ij}^2 V_i}\Big) = p\Big(E_i - \gamma_{ik}V_i/\sqrt{\beta\sum_i \gamma_{ik}^2 V_i}\Big)$. But the previous statement is only true when $\gamma_{ij}/\gamma_{ik} = \sqrt{\sum_i \gamma_{ij}^2 V_i}/\sqrt{\sum_i \gamma_{ik}^2 V_i}$. Now consider the point source profit maximizing payment to nonpoint source $l \neq i$. Since (2.8) holds for all nonpoint sources, it follows that $q_l^* = p\Big(E_l - \gamma_{lj}V_l/\sqrt{\beta\sum_l \gamma_{lj}^2 V_l}\Big) = p\Big(E_l - \gamma_{lk}V_l/\sqrt{\beta\sum_l \gamma_{lk}^2 V_l}\Big)$, which implies that $\gamma_{lj}/\gamma_{lk} = \sqrt{\sum_l \gamma_{lj}^2 V_l}/\sqrt{\sum_l \gamma_{lk}^2 V_l}$ must hold. Recognizing that $\sqrt{\sum_i \gamma_{ij}^2 V_i} \equiv \sqrt{\sum_l \gamma_{lj}^2 V_l}$ since the only difference is in the index, it follows that the following condition must hold if point sources j and k are to simultaneously maximize profits:

$$\frac{\gamma_{ij}}{\gamma_{ik}} = \frac{\gamma_{lj}}{\gamma_{lk}} = \frac{\sqrt{\sum_i \gamma_{ij}^2 V_i}}{\sqrt{\sum_i \gamma_{ik}^2 V_i}} \quad \forall\, i,j,k,l \in I \times J \tag{2.9}$$

Since the condition in (2.9) holds for any arbitrary pair of point sources it holds for all point sources. The first equality in (2.9) indicates that all pairs of point sources, who trade with nonpoint sources, buy equi-proportional shares in all nonpoint source credit-generating projects. If point source j buys twice as much abatement than point source k from nonpoint source i, then j will also buy twice as much abatement than

k from nonpoint source l. Point sources will spread the risk associated with nonpoint source credits by buying equal proportions (relatively speaking) in all nonpoint source credit generating projects. The second equality indicates that the ratio in which two point sources buy proportions is equal to the ratio between the total standard deviations of the point sources' total credit stock. Equation (2.9) indicates equi-proportional of sharing of relative risk among all pairs of point sources who trade with nonpoint sources.

It can also be shown that relative risk sharing implies risk sharing in the absolute sense among point sources. Using (2.9) I prove that each j buys the share γ_j in credit stock from all $i \in I$, i.e. for all $i \in I$ $\gamma_{ij} = \gamma_j$ for all $j \in J$. In words, when i sells its entire credit stock then j buys a γ_j share from it.

The proof follows. Consider a market with J point sources and I nonpoint sources. Let $\gamma_{lj} = b_j\gamma_{ij}$ and $\gamma_{lk} = b_k\gamma_{ik}$ for any quartet $i, j, k, l \in I \times J$. where i and l are the nonpoint sources and j and k are the point sources. The result in (2.9) indicates that $b_j = b_k = b$ for all $j, k \in J$. Since nonpoint sources sell their entire stock of abatement credits, for any $i, l \in I$ $\gamma_{ik} = 1 - \sum_{j \neq k} \gamma_{ij}$ and $\gamma_{lk} = 1 - \sum_{j \neq k} \gamma_{lj}$. But since $\gamma_{lj} = b\gamma_{ij}$ for all $j \in J$ it follows that $\gamma_{lk} = 1 - b\sum_{j \neq k} \gamma_{ij}$. Substituting for γ_{lj} and γ_{lk} in (2.9), the condition $(1 - \sum_j \gamma_{ij})/\gamma_{ij} = (1 - b\sum_j \gamma_{ij})/b\gamma_{ij}$ must hold to ensure cost minimization by all polluters. But the condition only holds when $b = 1$. Hence, it follows that $\gamma_{1j} = \cdots = \gamma_{ij} = \cdots = \gamma_{Ij} = \gamma_j$ for all $j \in J$. It also follows that $\sum_j \gamma_j = 1$. The point source j buys a γ_j share of credit stock from every nonpoint $i \in I$. The result may be stated as follows.

RESULT 1. *When the market has many traders and nonpoint sources sell their entire credit supply, then point sources control their exposure to risks from nonpoint source*

credits by buying identical shares in all nonpoint source credit-generating projects. This behavior may be described as Equi-Proportional Risk Sharing *across the different nonpoint credit stocks.*

Under Equi-Proportional Risk Sharing j's probabilistic allowance constraint (2.6) amends to $\sum_k \gamma_{kj} a_k + \gamma_j\big(\sum_i E_i - \sqrt{\sum_i V_i/\beta}\big) \geq 0$ through the substitution of γ_j for γ_{ij} but its profit function remains unchanged. Hence, the necessary conditions for profit maximization under equi-proportional risk sharing differs from the necessary conditions without equi-proportional risk sharing. The new necessary conditions are (2.7) and (2.10) below.

$$\sum_i \tilde{q}_i = p\left(\sum_i E_i - \sqrt{\frac{\sum_i V_i}{\beta}}\right) \tag{2.10}$$

where $\tilde{q}_i$ the payment to i for supplying $\tilde{a}_i(E_i, V_i)$ that maximizes j's profits under equi-proportional risk sharing. $\sum_i \tilde{q}_i$ is the total payment from all j to all i. As with q_i^*, the total payment $\sum_i \tilde{q}_i$ is additively separable into functions of E_i and V_i. By differentiating the RHS of (2.10) the optimal marginal payments associated with E_i and V_i are calculated. The marginal payment for E_i is p and the marginal payment for V_i is $-p/\sqrt{2\beta\sum_i V_i}$. Since p is also the marginal payment for e_j, E_i and e_j are treated equivalently at the profit maximizing margin. The credits supplied by i, $\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$, are rewarded for increasing E_i and penalized for increasing V_i. And by the definition of E_i and V_i, nonpoint sources are penalized for increasing the mean μ_i and variability σ_i^2 of their runoff. Under equi-proportional risk sharing point sources evaluate nonpoint source runoff in a manner identical to that required for the attainment of the safety-first environmental target at least cost [see (2.3)].

Note that the optimal relationship between $\tilde{q}_i$, E_i and V_i is not specified in (2.10). Theoretically, $\tilde{q}_i$ could be described by any function. Each nonpoint source could face a different payment function. The only condition is that they satisfy (2.10) in the aggregate. I propose a payment function of the form $\tilde{q}_i = p\Big(E_i(\mu_i) - V_i(\sigma_i^2)/\sqrt{\beta \sum_i V_i(\sigma_i^2)}\Big)$. The Cauchy-Schwarz inequality shows that summing over $\tilde{q}_i$ yields (2.10) and hence maximizes j's profits given equi-proportional risk sharing.[1] The results on the separability of $\sum_i \tilde{q}_i$ hold for $\tilde{q}_i$.

The payment function $\tilde{q}_i$ imposes conditions on the characteristics of the credits that i offers to the market. From the definition of $\tilde{q}_i$ it follows that i gets a positive payment if and only if $V_i/E_i \leq \sqrt{\beta \sum_i V_i}$ where V_i/E_i is the variance-to-mean ratio and $\sqrt{\beta \sum_i V_i}$ is standard deviation of all credits after adjusting for the allowable probability of individual constraint violation. The lower the allowable probability or the standard deviation of all credits, the lower must the variance of i's credits be relative to the mean. These conditions make intuitive sense. From (2.6), when β is reduced or when $\sum_i V_i$ falls, j's allowance constraint tightens. To satisfy this tightened constraint, j must reduce the risk in its credit portfolio, which it does by purchasing credits with low variance-to-mean ratios.

2.2.4 The Market Equilibrium

The necessary conditions for i's profit maximization amend given that it is paid $\tilde{q}_i$. Differentiating $\tilde{q}_i$ with respect to μ_i and σ_i^2, substituting into (2.4) and (2.5), and

[1] The Cauchy-Schwarz Inequality states that $\sum_i x_i y_i \leq \sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}$ with strict equality when $x_i = w \cdot y_i$ for all i and w is a scalar. Setting $x_i = \sqrt{V_i}$, $y_i = w \cdot \sqrt{V_i}$ and $w = 1$, the Cauchy-Schwarz Inequality indicates that $\sum_i V_i = \sqrt{\sum_i V_i}\sqrt{\sum_i V_i}$.

rearranging yields

$$\frac{\partial \pi_i}{\partial \mu_i} = \frac{\partial \pi_i}{\partial \sigma_i^2} \sqrt{\beta \sum_i \sigma_i^2} \left(\frac{2 \sum_i \sigma_i^2}{2 \sum_i \sigma_i^2 - \sigma_i^2} \right) = p \tag{2.11}$$

Equation (2.11) indicates that i maximizes profits by equating the marginal profits from increasing μ_i to the market price of point source credits p. The equation also supplies the necessary condition on σ_i^2 for i's profit maximization. In a competitive market $2 \sum_i \sigma_i^2 - \sigma_i^2 \approx 2 \sum_i \sigma_i^2$, in which case $\partial \pi_i / \partial \sigma_i^2 \approx -p / \sqrt{\beta \sum_i \sigma_i^2}$ where $\sqrt{\beta \sum_i \sigma_i^2}$ is the standard deviation of all emissions after adjusting for the allowable level of individual constraint violation. Since $\beta \sum_i V_i$ is identical across i, nonpoint sources simultaneously maximize profits by approximately equalizing marginal profits with respect to variance: $\partial \pi_i / \partial \sigma_i^2 \approx \partial \pi_l / \partial \sigma_l^2 \; \forall \; i, l \in I$.

Profit maximizing emissions variance σ_i^{2*} for an individual nonpoint source polluter is a function of aggregate emissions variance $\sum_i \sigma_i^2$. Since $\pi_i(\mu_i, \sigma_i^2)$ is concave in σ_i^2, when $\sum_i \sigma_i^2$ is high then i maximizes profits by setting σ_i^{2*} high and vice versa. The intuition might be as follows: Consider an nonpoint source that keeps σ_i^2 low when $\sum_i \sigma_i^2$ is high. By reducing σ_i^2 marginally i earns $p / \sqrt{\beta \sum_i \sigma_i^2}$, which is small when $\sum_i \sigma_i^2$ is high. The concavity of $\pi_i(\mu_i, \sigma_i^2)$ in σ_i^2 implies that the marginal reduction in σ_i^2 is costly when σ_i^2 is low. An assessment of costs and benefits would convince the nonpoint to keep σ_i^2 high when $\sum_i \sigma_i^2$ is high.

The market equilibrium is reached when all participants in the market have exhausted gains from trade and maximized their profits. This implies that the necessary conditions for profit maximization must hold for all i and for all j. In other words, at

the market equilibrium (2.7), (2.10) must hold for all j and (2.11) must hold for all i. Note that a necessary condition for (2.11) to hold is that (2.10) holds since the former was derived directly from the latter. Hence it is not necessary to mention (2.10) as an explicit condition for market equilibrium. The conditions for market equilibrium are formally stated in the Result below.

RESULT 2. *The MANA market is in equilibrium when the profit-maximizing condition* (2.7) *holds for all point sources and the profit-maximizing condition* (2.11) *holds for all nonpoint sources. The satisfaction of the latter condition is predicated upon each nonpoint source i being paid according to the payment function in* (2.10) *upon supply of the credit $\tilde{a}_i(E_i(\mu_i), V_i(\sigma_i^2))$. The former condition is predicated upon each point source j satisfying* (2.6).

2.2.5 The Least Cost Safety-First Outcome and the MANA Market Equilibrium

The pertinent issue for the Agency is whether conditions can be imposed on the market such that the safety-first environmental target (2.2) is met at least cost at the market equilibrium. This requires that the market equilibrium satisfy (2.3). A comparison of (2.3) and (2.11) reveals that in general the market equilibrium will not satisfy the safety-first environmental target at least cost because point sources evaluation of σ_i^2 differently from the Environmental Agency. Under assumption that there are many buyers and sellers, $2\sum_i \sigma_i^2 - \sigma_i^2 \approx 2\sum_i \sigma_i^2$, (2.3) and (2.11) indicate that $MRTS^{SF}_{\mu_i,\sigma_i^2}/MRTS^{ME}_{\mu_i,\sigma_i^2} \approx 2 \cdot (\alpha \sum_i \sigma_i^{2,SF}/\beta \sum_i \sigma_i^{2,ME})$. $MRTS$ is the marginal rate of technical substition and the SF and ME superscripts refer to least cost safety-first

solution and the market equilibrium respectively. In general $\alpha \sum_i \sigma_i^{2,SF} \neq \beta \sum_i \sigma_i^{2,ME}$. Further inferences on whether $\sigma_i^{2,ME}$ is greater than or less than $\sigma_i^{2,SF}$ cannot be drawn without parameterizing the abatement cost functions. It can be asserted, however, that without further restrictions the market equilibrium is not safety-first at least cost.

The divergence of the market equilibrium from least cost safety-first solution occurs because of the presence of β in the payment function $\tilde{q}_i$. By imposing appropriate conditions on β the market equilibrium will achieve the least cost safety-first outcome. Comparing (2.3) and (2.11) again, it is apparent that the nonpoint sources will choose values of σ_i^2 at market equilibrium that lead to the least cost safety-first outcome when $\sqrt{\beta/\alpha} = 2 - \sigma_i^2 / \sum_i \sigma_i^2$. In a competitive market $\sigma_i^2 / \sum_i \sigma_i^2 \approx 0$. Hence, the MANA market equilibrium satisfies the safety-first environmental outcome at least cost when $\sqrt{\beta/\alpha} \approx 2$ and $\beta \approx 4\alpha$. The allowable probability that individual point sources violate their allowances should be about four times greater than the allowable probability of that the safety-first environmental target is violated. If the acceptable probability at the aggregate level is 5% then individual point sources should be allowed a 20% probability of violation.

Result 3. *The equilibrium in the MANA market satisfies the least cost safety-first outcome under a single condition: that the acceptable probability with which point sources may violate their allowance, β, be chosen such that $\beta \approx 4\alpha$ where α is the acceptable probability with which the safety-first environmental target may be violated.*

Result 3 indicates that it is trivial to ensure that the equilibrium outcome in the market with multi-attribute nonpoint source credits is safety-first at least cost since β is trivially calculated once α is known.

2.3 The Trading Ratio Market

Extant water quality trading markets do not use a multi-attribute definition of the nonpoint source credit. Neither do they impose individual point source loadings constraints based on safety-first environmental targets. Instead the nonpoint source credit in expected terms. For nonpoint source i the credit is simply defined as E_i, which is a first order approximation of the true reduction in emissions below the allowance, $a_i = \hat{r}_i - r_i$. Using a first order approximation prevents transmission of information about the risk or variability σ_i^2 associated with i's runoff to the marketplace. The variance σ_i^2 will not affect trading decisions. However, σ_i^2 is a factor in the satisfaction of the safety-first environmental target (2.2) and it must be controlled if (2.2) is to be satisfied at least cost.

Instead of using the market mechanism to control σ_i^2, an *ad hoc* adjustment called the 'trading ratio' is used to control emissions risk in current water quality trading markets. The trading ratio or TR is defined as the number of units of mean nonpoint source abatements required to allow a unit increase in point source loadings (Shortle 1990). It is calculated by the Environmental Agency and applied to all trades where nonpoint sources participate. Proponents of the TR argue that when it is correctly designed it controls the stochasticity of nonpoint emissions. Through its application

point and nonpoint source credits are converted into perfect substitutes so that in effect apples will trade for apples (Woodward 2000).

There is no consensus on the design of the optimal TR. For example, Horan and Shortle (2005) find that the optimal TR must reflect the marginal environmental impacts of the trading polluters, the riskiness of their abatements and the transactions costs of the trade. In contrast Malik et al. (1993) find that the optimal TR must also reflect abatement and enforcement costs. In spite of the differences the general result is that the optimal TR will differ by nonpoint source. I analyze whether the TR can be used to achieve the safety-first environmental target at least cost.

2.3.1 The Trading Ratio Market Equilibrium

As in the MANA market, the point source j earns revenues from sale of credits to other point sources, and faces costs from credit generation and purchase of credits from nonpoint and other point sources. The point source j's profit function is $\Pi_j = p[a_j - \sum_k \gamma_{kj} a_k] - [\pi_i^0 - \pi_j(a_j)] - \sum_i \gamma_{ij} q_i(E_i(\mu_i))$ where $p[a_j - \sum_k \gamma_{kj} a_k]$ is net income from trade with other point sources, $[\pi_i^0 - \pi_j(a_j)]$ is the cost of credit generation and $\sum_i \gamma_{ij} q_i(E_i(\mu_i))$ is the payment to nonpoint sources for procurement of $\sum_i \gamma_{ij} E_i(\mu_i)$ credits. Let t_{ij} be the trading ratio applied to a trade between nonpoint source i and point source j. If $t_{ij} = 2$ then by buying two units of abatement from i, entitles j to increase its emissions by one unit. The emissions allowance for j is $\sum_k \gamma_{kj} a_k + \sum_i (\gamma_{ij}/t_{ij}) E_i(\mu_i) \geq 0$. The necessary conditions for j to minimize costs subject to its allowance are (2.7) and (2.12).

$$q_i^t = p \frac{E_i(\mu_i)}{t_{ij}} \quad \forall\, i \in I \tag{2.12}$$

The necessary conditions are obtained from manipulation of the Kuhn-Tucker conditions. The point source maximizes its profits if nonpoint source i is paid q_i^t for supplying $\gamma_{ij}E_i$ credits. The variance of nonpoint source emissions is not a pricing factor in the TR-based market. Equation (2.12) indicates that payments to i decrease when t_{ij} increases. Also, since $E_i = \hat{r}_i - \mu_i$ it follows that q_i^t is inversely related to μ_i. Nonpoint sources with high mean emissions are paid less, which is consistent with a policy geared towards reducing mean emissions.

Since (2.12) is a necessary profit maximizing condition for all j it follows that $q_i^t = p(E_i/t_{ij}) = p(E_i/t_{ik})$ for all $j, k \in J$, which in turn implies that $t_{ij} = t_{ik} = t_i$ for all $j, k \in J$. The trading ratio should be set such that it varies across nonpoint sources but not across individual trades. This result differs from previous results in the literature where the trading ratio is trade-specific, affected as it is by point source and nonpoint source characteristics (Malik et al. 1993; Horan and Shortle 2005). The differences stem from the modeling assumptions and the fact that earlier models only studied interaction between a single point source and a single nonpoint source. The entire market was not modeled.

Let us now look at the nonpoint source profit maximization problem in the TR-based market. Like in the MANA market nonpoint source i has one cost and one revenue stream. The cost is that of generating the mean revenue and is $\pi_i^0 - \pi_(\mu_i)$, where π_i^0 is the unconstrained maximized profit and μ_i is mean emissions associated with the nonpoint source credit. As defined in Section 2.2.1 $E_i = \hat{r}_i - \mu_i$ where $\hat{r}_i$ is i's allowance. The revenue, $q_i^t(E_i) = pE_i/t_i$ comes from the sale of credits. Nonpoint source i's total profits are $\Pi_i = pE_i(\mu_i)/t_i - [\pi_i^0 - \pi_(\mu_i)]$. Like in the MANA market the i does not face a

binding constraint. It optimizes by choosing μ_i such that profits are maximized. The necessary condition for i to maximize profits is

$$\frac{\partial \pi_i}{\partial \mu_i} = \frac{p}{t_i} \tag{2.13}$$

The necessary condition implies that i can only maximize profits if the Environmental Agency chooses the TR, $\tilde{t}_i$, such that $\tilde{t}_i = p \cdot \partial \mu_i / \partial \pi_i$. Since $\partial \pi_i / \partial \mu_i \geq 0$ it follows that $\tilde{t}_i \geq 0$. The concavity of $\pi_i(\mu_i)$ indicates a directly proportional relationship between $\tilde{t}_i$ and μ_i. When μ_i increases marginally then $\partial \pi_i / \partial \mu_i$ decreases marginally (by concavity), which implies that $\tilde{t}_i$ increases marginally (by (2.13)).

Substituting $\tilde{t}_i$ into (2.12) yields the payment that maximizes i's profits, $\tilde{q}_i^t = (\hat{r}_i - \mu_i)\partial \pi_i / \partial \mu_i$. Since (2.12) is a condition for point source profit maximization, $\tilde{q}_i^t$ maximizes both point and nonpoint source profits. Note that $\tilde{q}_i^t \to 0$ as $\mu_i \to \hat{r}_i$ and $\tilde{q}_i^t \to \hat{r}_i \partial \pi_i / \partial \mu_i$ as $\mu_i \to 0$. The concavity of π_i with respect to μ_i implies that it is likely that $\partial \pi_i / \partial \mu_i > 1$ when $\mu \to 0$. This in turn implies that $\tilde{q}_i^t$ is more that proportionately responsive to changes in μ_i. The price of E_i increases more quickly than the size of E_i does: if one unit of E_i cost \$ x then two units would cost more than \$ $2x$, which implies strong market preferences for credits with low μ_i.

At the market equilibrium all polluters are profit maximizers, which implies that (2.7) and (2.13) must hold simultaneously for all i and j. The necessary condition for market equilibrium in the trading ratio market is

$$\frac{\partial \pi_j}{\partial e_j} = \tilde{t}_i \frac{\partial \pi_i}{\partial \mu_i} \quad \forall\, i, j \in I \times J \tag{2.14}$$

At the TR-based market equilibrium marginal costs of point and nonpoint credits should be equated after adjusting for the trading ratio. The results are similar to the equi-marginality results that characterize the standard pollution trading models, except for the role played by the trading ratio.

Note that the optimal choice of the TR requires that the Environmental Agency know the market equilibrium outcome which, in essence, implies perfect information. However, if this information were available then a market for water quality trading is redundant because the Agency could directly allocate allowances such that its safety-first target is met at least cost.

2.3.2 The TR-Based Market Equilibrium and the Safety-First Environmental Target

The least cost safety-first outcome requires control of μ_i and σ_i^2 for all i and e_j for all j, implying control of $2I+J$ variables. In the TR-based market equilibrium only $I+J$ variables are controlled because the I σ_i^2s are irrelevant to the solution. Except when the Environmental Agency has perfect information, the TR-based market will never achieve the least cost safety-first outcome because it allows control over too few variables of interest. The only scenario where the TR-based market equilibrium does achieve the least cost safety-first outcome is when the latter outcome is achieved by controlling $I+J$ variables instead of $2I+J$ variables. The latter outcome is feasible when I variables in the Agency's problem are deterministic functions of the other variables in the Agency's problem. Since nonpoint abatements are independent this is only possible when σ_i^2 is a deterministic function of μ_i or $\sigma_i^2 = \sigma_i^2(\mu_i)$ for all $i \in I$. In this scenario controlling μ_i

controls σ_i^2. Given this scenario the least cost safety-first outcome transforms from (2.3) to

$$\frac{\partial \pi_j}{\partial e_j} = \frac{\partial \pi_i}{\partial \mu_i} \frac{2\sqrt{\alpha \sum_i \sigma_i^2(\mu_i)}}{2\sqrt{\alpha \sum_i \sigma_i^2(\mu_i)} + \partial \sigma_i^2 / \partial \mu_i} \tag{2.15}$$

The TR-based market will satisfy the safety-first environmental target at least cost when (2.14) and (2.15) are identical, which requires the following necessary condition:

$$\frac{1}{\tilde{t}_i} = 1 + \frac{1}{2\sqrt{\alpha \sum_i \sigma_i^2(\mu_i)}} \frac{\partial \sigma_i^2}{\partial \mu_i} \quad \forall\, i \in I \tag{2.16}$$

When $\partial \sigma_i^2 / \partial \mu_i > 0$ then $\tilde{t}_i < 1$ and the optimal μ_i will be relatively high given the concavity of $\pi_i(\mu_i)$ with respect to μ_i. Conversely when $\partial \sigma_i^2 / \partial \mu_i < 0$ then $\tilde{t}_i > 1$ and the optimal μ_i will be relatively low. The safety-first least cost outcome requires the TR chosen such that nonpoint sources are penalized when σ_i^2 and μ_i trend in opposite directions and rewarded when they move in the same direction. When μ_i and σ_i^2 trend in the same direction then trade in credits based on emissions with high μ_i and high σ_i^2 is optimal from the safety-first point of view. A low TR will encourage the generation of and trade in such credits. Analogously when μ_i and σ_i^2 trend in opposite directions then trade in credits based on emissions with low μ_i and high σ_i^2 is optimal from the safety-first point of view. A high TR will encourage such trade.

The TR-based market always fails when μ_i and σ_i^2 are independent. When $\partial \mu_i / \partial \sigma_i^2 \approx 0$ then the TR-based market is a poor instrument for the control of σ_i^2. In this case the TR at market equilibrium is $\tilde{t}_i \approx 1$ for all i and the marginal costs of (expected) abatement are equated across all polluters. The Agency will be unable

to ensure attainment of the safety-first environmental target at least cost through the TR-based market. However, whatever the aggregate pollution level may be, it will be achieved at least cost.

RESULT 4. *Equilibrium in the trading ratio market satisfies the safety-first environmental target at least cost under three conditions. First, the Environmental Agency must have perfect information because it must have knowledge of the market equilibrium to optimally allocate trading ratios. Second, the variance of all nonpoint emissions must be deterministic functions of the mean:* $\sigma_i^2 = \sigma_i^2(\mu_i)\ \forall\ i \in I$. *Finally, the relationship between the trading ratio* $\tilde{t}_i$, σ_i^2 *and* μ_i *must satisfy* (2.16).

In most existing markets $\tilde{t}_i = 2$ and uniform (Breetz et al. 2004; Morgan and Wolverton 2005). From (2.16), a uniform trading ratio of two is optimal from the least cost safety-first point of view only if $-\partial\sigma_i^2/\partial\mu_i = \sqrt{\alpha \sum_i \sigma_i^2}$ for all i. When $\sum_i \sigma_i^2$ is large then optimality requires that σ_i^2 be *extremely* responsive to changes in μ_i. Determination of whether such responsiveness is feasible, indeed possible, is beyond the scope of this paper, but it does seem unlikely. Instead, if high responsiveness of σ_i^2 to μ_i is infeasible, (2.16) indicates that $\tilde{t}_i$ will be closer to one than two. The result works for when $\partial\sigma_i^2/\partial\mu_i > 0$ too. A trading ratio of 2/3 would imply that $\partial\sigma_i^2/\partial\mu_i = \sqrt{\alpha \sum_i \sigma_i^2}$. Variance of emissions will have to be extremely responsive to the mean for a trading ratio of 2/3 to be optimal. If such responsiveness is infeasible then $\tilde{t}_i$ will be closer to one than 2/3.

In a scenario where allowances are externally chosen, if the trading ratio market equilibrium is to satisfy the safety-first market outcome at least cost, then $\tilde{t}_i$ is close to one

for all i. The uniform trading ratio of two, which is commonly used in existing markets, is sub-optimal. These results imply that trading ratios are very sensitive instruments for achieving safety-first environmental targets and must be calibrated very precisely. But such precision is unlikely in a stochastic environment where the true relationship between μ_i and σ_i^2 are unknown. In the real world the assumed relationship is merely a working hypothesis. I infer is that the trading ratio is a poor instrument for attaining the safety-first environmental target at least cost.

2.4 Discussion and Conclusions

The equilibrium in the TR-based market satisfies the safety-first target at least cost when three conditions hold. First, the variance of the nonpoint emissions must be a deterministic function of the mean since variance is not directly controlled in the TR-based market, but control is necessary for the attainment of the least cost safety-first outcome. The second condition specifies the optimal form of the trading ratio (2.16). It imposes strict constraints on the relationship between μ_i and σ_i^2. The final condition is that the Agency have perfect information and knowledge of the market equilibrium, necessary for the optimal calculation of $\tilde{t}_i$.

It is unlikely that any of these conditions can be satisfied in a real world trading environment. The Agency does not have information on the private costs of polluters and at the current state of the science the probability distribution of nonpoint runoff is unknown. Instead, through observation and modeling an empirical approximation of the true probability distribution is estimated. Imposing a deterministic relationship between μ_i and σ_i^2 requires that a parametric distribution be fitted to the data. The

validity is such an imposition cannot be ascertained. A wide variety of distributions can typically be fitted to the data, each specifying different marginal relationships between μ_i and σ_i^2. The choice between these distributions is arbitrary, which implies that the assumed relationship between μ_i and σ_i^2 is arbitrarily assigned, which in turn implies that the choice of trading ratio is arbitrary. Precise calibration of the trading ratio is impossible in the real world. But as discussed in Section 2.3.2 the trading ratio market is an extremely sensitive market instrument and fails if trading ratios are not correctly calibrated. In the real world the trading ratio market will not achieve the least cost safety-first outcome.

The MANA market does not require a deterministic relationship between E_i and V_i and does not require sensitive calibration. Instead it relies on two rules and the market mechanism to generate the least cost safety-first outcome. The first rule defines the tradable nonpoint source credit in terms of μ_i and σ_i^2. The second is that the allowable probability with which a point source violates its allowance should be approximately four times greater than the allowable probability with which the safety-first environmental target is violated. When these rules are satisfied then the market equilibrium always satisfies the safety-first target at least cost.

The degree of regulator intervention is also much lower in the market with multi-attribute goods. The regulator's role is purely informational: it is required to provide information on μ_i and σ_i^2 for all nonpoint source credits. In the trading ratio market, by contrast, apart from providing information on μ_i the regulator is also required to calculate the optimal trading ratio $\tilde{t}_i$ for all nonpoint sources. Although not modeled, transactions costs will be higher in real world trading ratio markets.

Numerous extensions to the model are possible. An interesting possibility is to study the market equilibria when nonpoint source loadings are correlated, which is a realistic assumption for proximal nonpoint sources that have similar topography and weather. Another possibility is to conduct a series of economic experiments to test the robustness of the results to violations in underlying assumptions. The assumption of many buyers and sellers might be removed. Real world markets are oligopsonistic with few large point sources and many small nonpoint sources. Another removable assumption pertains to the continuity of abatement production function. In reality decisions have discrete consequences and are not made using the marginality analysis that underpins our results. The experiments will be relevant to policy design because they represent an intermediate step in the transfer of policy ideas from the academic to the real world.

Chapter 3

An Experimental Comparison of Two Markets for Controlling Nonpoint Source Pollution

Markets in tradable property rights are commonly regarded with approbation as efficient mechanisms for the attainment of specific policy targets at minimum social cost (or maximum social benefit) in the presence of externalities.[1] There has been much application of economic theory and experiments to the design of markets to correct for the distortional impacts of environmental externalities. A rich literature shows that the imposition of an enforceable property rights regime with provision to trade will result in significant amelioration of the environmental problem. In the pollution context, tradable property rights are typically characterized as pollution licenses, emissions permits or abatement credits, which bestow upon the owner the right to discharge a specified quantity of pollutants (Hahn and Hester 1989; Montgomery 1972).

3.0.1 Early Emissions Markets Experiments

The optimal design and administration of markets in tradable property rights when used as incentive-based mechanisms for the decentralized control of pollution has been a research objective among experimental economists since the early 1980s. The early research focused on the control of pollution from point sources. We draw on Muller and Mestelman (1998) for some of the subsequent discussion on early emissions trading

[1] Externalities manifest whenever the utility of a consumer or production possibilities of a producer are affected by the activity of another economic agent (Mas-Colell et al. 1995, pg. 352)

research. Laboratory research into emissions trading began with Plott (1983), whose results supported permit trading as a feasible mechanism for controlling pollution externalities. Other early areas of interest were market power, revenue neutrality, banking and evaluation of existing point source trading programs like EPA's Acid Rain Control Program (Cason and Plott 1996), and the nitrous oxide trading program in southern Ontario (Muller and Mestelman 1994).

One controversy in the debate over tradable permit markets was over its role as a government revenue stream. Dales (1968) envisioned the markets as a revenue source: the revenues would be generated through an auction of initial permit holdings. Citing political feasibility, other researchers proposed free permit allocations to polluters and a revenue neutral role for emissions trading (Hahn and Noll 1982; Hahn 1988). A series of experiments analyzed the performance of Hahn and Noll's Revenue Neutral Auction (RNA). Franciosi et al. (1993) compare the RNA to the Uniform Price Auction – a revenue generating mechanism – and find no significant differences between the two institutions. Ledyard and Szakaly-Moore (1994) compare the RNA to a revenue neutral double auction and find evidence in favor of the double auction. Subsequent research has largely discarded the RNA, focusing instead on auction (e.g., Cason and Gangadharan 2003) or market mechanisms (e.g., Ben-David et al. 1999) with no explicit cognizance of the role of revenue neutrality.

Other early research focused on the banking of unused permits for use in subsequent periods. Cronshaw and Brown-Kruse (1996) compare a banking-only environment to an alternate that allows banking *and* trading and find that the banking and trading environment performs unambiguously better. Other studies compared trading-only

environments to banking-and-trading environments and found significant levels of underbanking (Franciosi et al. 1992; Godby et al. 1997; Cason et al. 1999). The implication is that banking did not increase market efficiency, perhaps because of added complexity in the decision environment and introduction of the possibility of inter-temporal speculation. Banking does however seem to reduce credit price variability in uncertain environments (Carlson and Sholtz 1994; Godby et al. 1997; Cason and Gangadharan 2006).

Market power through size and information asymmetry is always a concern when designing markets and there is ample evidence of this concern in the emissions trading market literature. Experimental results are mixed. Most studies do confirm the presence of market power in asymmetric markets – with few buyers and many sellers or vice versa – but to a lesser degree than indicated by the theory (Muller et al. 2002; Cason et al. 2003). A minority find no evidence of market power, with prices and trades converging to the competitive level (e.g., Carlèn 2003). We note that the cited studies were conducted using a continuous double auction environment, which invites speculation on how discrete trading environments might impact market power.

3.0.2 Experiments in Nonpoint Pollution Control

Perhaps as a consequence of the development and maturing of the experimental literature on emissions trading and pollution control, recent years have seen growth in research related to the control of nonpoint source pollution and point-nonpoint trading. Nonpoint source pollution is diffuse, which raises considerations that are extraneous to research and policy on point source pollution control. The primary consideration is

that its diffused nature precludes accurate and inexpensive metering and monitoring (Carpentier et al. 1998; Harrington et al. 1985). Instead, pollution from a particular nonpoint source is typically estimated through simulation. The simulation approach generates unreliable estimates for two reasons. First, the simulations are handicapped by a lack of validated data on the relationship between polluter and downstream impact (DeCoursey 1985; Negahban et al. 1994). Second, agricultural loadings are inherently stochastic because they are affected by weather-related and other random factors (Shortle and Dunn 1986).

There are two distinct bodies of theoretical literature on the control of nonpoint source pollution. One focuses on the design and use of emissions trading markets analogous to those used to control point source polluters (e.g., Horan and Shortle 2005; Hung and Shaw 2005). The other focuses on collective mechanisms of reward and punishment like ambient taxes and group contracts (e.g., Segerson 1988; Cabe and Herriges 1992).

A sizable experimental literature exists on the performance of collective non-market mechanisms. The evidence is mixed for ambient tax / subsidy instruments with some results indicating efficiency (Spraggon 2002) and others indicating inefficiency through overabatement (Cochard et al. 2005). The over-abatement increases when subjects communicate (Vossler et al. 2006; Poe et al. 2004). Group fines are found to be less efficient and less reliable (Cochard et al. 2005; Spraggon 2002), but efficiency increases leading to optimal compliance when cheap talk is permitted (Vossler et al. 2006). The performance of ambient instruments is also affected by polluter heterogeneity. Providing support to contentions by Shortle and Horan (2001) and Weersink et al. (1998)

that ambient instruments are only appropriate when polluters are small and homogenous, Spraggon (2004) finds inefficiencies caused by freeriding by small polluters at the expense of large polluters. However, the freeriding can be controlled when a reverse auction to uncover hidden polluter information is used in conjunction with group contracts (Taylor et al. 2004).

Despite the academic interest, ambient instruments face political limitations (Shortle and Abler 1994; Xepapadeas 1999), which makes their implementation infeasible. Ambient instruments are vulnerable to freeriding because the natural variability in emissions makes shirking a viable behavior (Shortle and Horan 2001). The policy community has instead focused on tradable permit markets as the choice economic instrument for nonpoint pollution control, as evidenced by the initiatives taken by the USEPA and state environmental boards (Breetz et al. 2004). The expectation is that these markets can result in cost savings in excess of US$ 1 billion when compared to command-and-control mechanisms (USEPA 2001).

The literature proposes a two-step adjustment to the "textbook" model of trading between point sources to account for the variability and non-measurability of nonpoint abatement. First, nonpoint abatement is defined by its mean and then converted into credits. Second, the mean-based nonpoint credit is exchanged with point source credits subject to a trading ratio, defined as the number of units of nonpoint credits that must be purchased to allow a unit increase in point source pollution (Shortle 1990). The trading ratio when correctly calculated will account for the marginal damages and uncertainty of each nonpoint abatement and the transaction costs associated with each

trade (Horan and Shortle 2005; Malik et al. 1993). Correct calculation requires knowledge of the environmental damage function and the emissions demand of all polluters. Insufficient development of the science and information asymmetries hamper access to this knowledge.

Practically, optimal calculation of the trading ratio is infeasible, which impedes the performance of trading ratio based markets (Woodward 2000). Ghosh and Shortle (2008) suggest that such markets will only be efficient under strong conditions on the relationship between the mean and variability of abatement. Validation of the existence of these conditions through data is problematic. We also point out that in existing markets nonpoint abatement is predicated upon the implementation of specific technologies. The technologies are separable and distinct, implying discreteness in the abatement production function. This discreteness brings into question the theoretical inferences on point-nonpoint trading markets, which assumes continuity in abatement production.

We were unable to find experimental validation of trading ratio-based markets as instruments of nonpoint pollution control, but there are studies focusing on specific aspects of the problem. Cason and Gangadharan (2006) tackle emissions variability through a test bed where emissions are variable but measurable and banking is allowed. They find that banking mitigates credit price variability as a consequence of the shocks, but at the expense of non-compliance and higher emissions levels. Ben-David et al. (1999) study the impact of technological heterogeneity among firms on permit market performance. Contrary to theoretical expectations they find reduced trade volumes and decreased market efficiency.

3.0.3 Elements of Our Experiment

We attempt to fill a gap in the experimental literature by focusing our analysis on emissions trading markets for point and nonpoint sources. We compare two market institutions[2] for credit trading in an environment that contains three features common to existing markets for point-nonpoint emissions trading: market power, abatement stochasticity and the discreteness of abatement technologies. These features, when in isolation, imply deviations from market efficient outcomes, as validated by the theory and previous experimental work. Their collective impact on market performance is unknown. A formal analysis is difficult because of mathematical intractability. Experimental analyses do not exist. Yet, analysis of their collective impact has real world relevance. Existing point-nonpoint trading markets exhibit all three characteristics. By comparing the status quo market institution for nonpoint pollution control to a theoretically superior alternative in an environment that captures three important elements of real world markets, we hope to generate inferences on market design that are directly applicable to the real world.

The two market institutions compared in this experiment differ in their treatment of nonpoint abatement stochasticity. The first institution is based on the status quo trading ratio market and has similar rules. Hereafter referred to as the TR market, in this institution nonpoint abatement is defined by its mean and traded subject to the trading ratio. In the second institution the nonpoint abatement credit is a two-attribute good, where the two attributes are the mean and variance of the underlying abatement.

[2]Market institutions may be thought of as sets of individual property rules under which agents communicate and exchange commodities to modify their initial endowments in accordance with their preferences (Smith 1982).

We call this market the Multi-Attribute Nonpoint Abatement or MANA market. A formal analysis of this market is available in Ghosh and Shortle (2008). The rationale behind the broader definition of abatement is that it supplies information on abatement risk to the market. Buyers and sellers are able to explicitly price risk when choosing their credit portfolios, unlike in trading ratio markets where the regulator corrects for risk through intervention (by trading ratio) in all trades.

We incorporate market power into the experiment by assuming that the emissions market is characterized by oligopsony and size asymmetry. There are few buyers of credits and many sellers. Furthermore the buyers are much larger economic entities. This structure mimics that of most existing point-nonpoint trading programs such those on the Kalamazoo river, in the Tar-Pamlico basin and on the Rock river (Morgan and Wolverton 2005). Size asymmetry has often been analyzed experimentally in point source only emissions markets. Its incorporation into a point-nonpoint setting is a design innovation.

We assume that point sources only buy credits and nonpoint sources only sell credits. In existing markets nonpoint sources have no incentive to buy credits because they do not face mandatory emissions caps. Rather, the expectation is that emissions markets will control nonpoint emission *without* direct regulation, through financial incentives provided by point sources (USEPA 2003). We do not allow trades between point sources because such trades will obscure point-nonpoint interactions, which are the phenomena of interest.

The final inefficiency causing feature in the environment is that nonpoint abatement is technologically discrete and heterogeneous. This is an understudied issue in both the theoretical and experimental literatures. The discreteness stems from the manner

in which nonpoint abatement credits are calculated in trading programs like the Chesapeake Bay program (for rules see PADEP 2006). Since abatement is not measurable it is predicated on abatement technologies implemented by the nonpoint source after accounting for local characteristics. Hence if a nonpoint implements technology A he is awarded a credits, if he implements B he is awarded b credits and so on. The technologies serve as emissions proxies because they are correlated with emissions.

Under easily implemented rules, no market power and continuity in the production of abatement, Ghosh and Shortle (2008) show that the market equilibrium in the MANA market will be always be socially optimal, and market outcomes in the TR market will always be sub-optimal except under extremely restrictive conditions. In this paper we test the robustness of those results in experimental settings that approximate important aspects of real world markets. Importantly, we allow for market power and a discrete decision making environment.

3.1 Experimental Design

We conduct eight sessions in which ten subjects trade credits derived from nonpoint abatements in a computerized uniform price call market trading system.[3] Instructions and the visual environment made use of neutral terminology to remove the effect of subjective environmental biases on decision making. For example credits were called

[3]The call market environment is preferred to a continuous double auction environment for many reasons. First, a call market is more efficient than a double auction when trading volumes are low (Economides and Schwartz 1995). Second, it eliminates buy-ask spreads and so removes a significant portion of the price variability that exists in double auctions (Economides and Schwartz 1995). Third, its discrete nature implies lower administrative costs than in the continuous double auction environment (Cason and Plott 1996). Existing point-nonpoint credit trading markets have low trading volumes (see Morgan and Wolverton 2005).

"fips" and abatement creating technologies were referred to as "projects." All subjects were graduate and undergraduate students at Pennsylvania State University and the sessions were conducted between November 17, 2008 and April 22, 2009.

3.1.1 Subjects

Two subjects are cast as point sources buying credits and the remaining eight are cast as nonpoint sources selling credits when the experimental session begins. Subjects keep their job roles for the duration of the session. Buyers cannot sell credits and sellers cannot buy credits. The description and distribution of roles reflects the oligopsony and incentive structure in current point-nonpoint markets. For the remainder of this paper we use "point source" and "buyer" interchangeably and "nonpoint source" and "seller" interchangeably. The market consists of three types of subjects: two seller types and one buyer type. Subjects within a type are homogeneous.

Sellers are divided into two groups of four and are homogeneous within groups but heterogeneous across groups. Each seller type has access to three technologies that supply stochastic abatements. The seller is aware of the credits supplied by each technology and its implementation cost. In any round she may offer only one of the three projects to the market. By imposing this rule we abstract from interactions between technologies and their impact on cost and credit generation. Also, the rule is useful as a mechanism for increasing seller participation because it increases the likelihood that individual sellers are able to trade successfully. Cason and Gangadharan (2005) made a similar observation in the context of credit auctions.

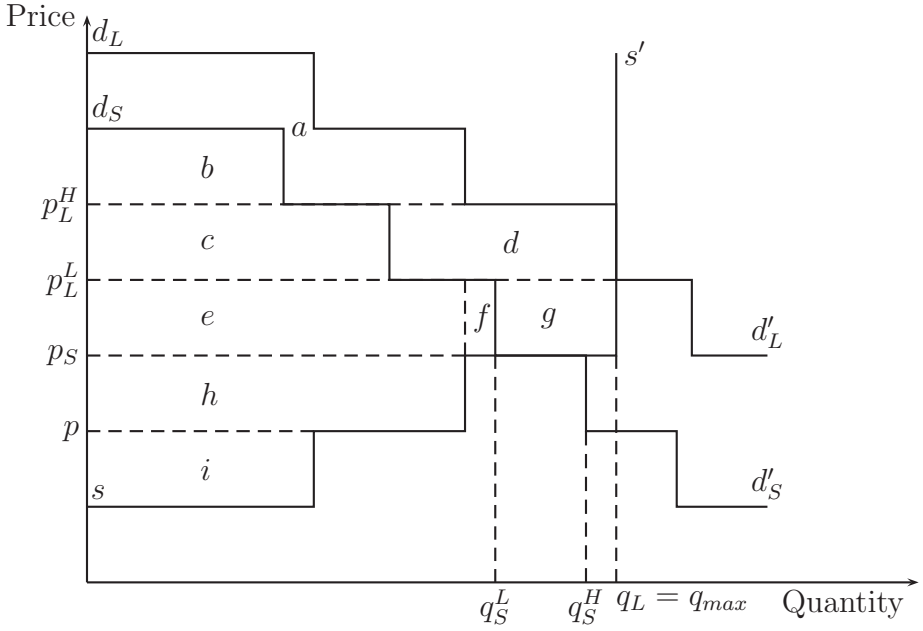
Buyers are homogeneous. Each buyer faces an emissions cap, which he meets by credit purchase and own abatement. He buys credits by contracting with a seller to implement a given project. His credit stock increases by the quantity of credits supplied by the project. Buying credits is profitable because it allows expansion of production. This profit is characterized in the experiment as "redemption value." Typically the more credits supplied by a project the higher its redemption value. We assume that the marginal benefits from credit purchase are decreasing. In every round a buyer can make multiple bids for projects. Since marginal benefits decrease in credits bought, redemption values associated with later bids are less than those associated with earlier bids: a project with a redemption value of \$70 when the market opens might only have a redemption value of \$50 conditional on ownership of two other projects. Since there are two seller types, each with access to three technologies, there are six projects potentially on offer in the market.

3.1.2 Treatments

There are two interacted treatments in this experiment. The first treatment uses a between-subjects design to study the impact of size on market power. In half the sessions the buyers are large and aggregate credit supply is exhausted at the competitive equilibrium. In the other sessions buyers are smaller, aggregate demand is lower and credits remain unsold at the competitive equilibrium. The two cases are shown in Figure 3.1 where ss' represents aggregate supply, $d_L d'_L$ is aggregate demand when buyers are large and $d_S d'_S$ is aggregate demand when buyers are small. Supply increases with

price until price reaches p_S, at which point credit supply reaches its maximum q_{max}. Thereafter credit supply is perfectly inelastic.

Fig. 3.1. Size and Market Power Treatment



The second treatment contrasts the MANA and TR market institutions. The markets are compared through a within-subjects crossover design where the two market types are alternated over successive periods. There are four MANA markets and four TR markets every session. Each MANA market shares its parameter set with a TR market implying four MANA-TR market pairs per session.

The two market institutions differ only in the way abatement credits are defined. In the MANA market the number of credits (or fips) supplied by a project is simply $q_S^M = E - V$, where E is the expected value and V the variance of abatement. We capture, in the simplest way possible, the idea driving the MANA market: that when

pollution generation is stochastic, market outcomes are socially optimal only if credits are defined to reflect preference for abatements with high E and low V (see Ghosh and Shortle 2008). q_S^M may be interpreted as a measure of project quality: the greater the quantity of q_S^M supplied by a project, the higher its quality. A project that supplies an abatement with $E = 12$ and $V = 9$ generates $q_S^M = 3$.

In the TR market the credits supplied by a project is $q_S^T = E/t$ where t is the project-specific trading ratio. If the project with $E = 12$ and $V = 9$ has $t = 2$ it generates $q_S^T = 6$. Trading ratios are allocated in a similar manner to that in existing point-nonpoint trading programs. In half the periods a uniform trading ratio of 2:1 or 3:1 is used, reflecting arrangements in the Tar-Pamlico program, New York City Watershed Offsets Pilot Program and others. In the other periods there is limited differentiation by nonpoint type and project like in programs on the Kalamazoo, Charles and Rock rivers (for details see Breetz et al. 2004).

The complex trading ratios found in the theory (see Hung and Shaw 2005; Malik et al. 1993; Horan and Shortle 2005) are not used for three reasons. First, the marginality conditions that define the optimal trading ratio are meaningless in a discrete trading environment. Second, reflecting ground realities in existing markets, we assume that the environmental damage function is unknown. Third, the models in the cited papers assume a 1×1 trading environment and the transferability of those results to a more complex environment is unknown. Ghosh and Shortle (2008) show that in a cap-and-trade market with many players trading ratios are easier to compute and depend only on seller characteristics and the price of point source credits. Information asymmetries that hinder calculation of even these simplified trading ratios will remain.

We implement a balanced 2x2 design with the buyer size treatment implemented through a between-subjects approach and the market institution treatment implemented through a within-subjects approach. There were four sessions with large buyers and four with small buyers. In each session there were four MANA markets and four TR markets. The treatments are distributed across 64 market periods of data as shown in Table 3.1.

Table 3.1. Distribution of Treatments across Periods

	TR Market	MANA Market	Total
Large Buyers	16	16	32
Small Buyers	16	16	32
Total	32	32	64

3.1.3 Sessions

Each experimental session consists of eight periods. Each period consists of a minimum of three and maximum of eight rounds. Periods end either when the credit price has converged or if eight rounds have passed. We assume convergence when prices remains unchanged across two periods. The rounds-within-periods format is common to market experiments (e.g., Bossaerts et al. 2002), although most market experiments on emissions trading have designs where each period consists of a single round (e.g., Cason and Plott 1996; Franciosi et al. 1993; Muller et al. 2002). We find the rounds-within-periods approach more suited to our experimental framework, where the markets in each period are structurally identical but have different parameterizations resulting in different abatements and costs. The rounds-within-periods approach allows us to track

price evolution in changing cost and production environments and get hence deeper insight into the price determination process.

Each session has four parameter sets. Since there are two types of sessions – for small and large buyers – eight parameter sets were used in this experiment. Table 3.2 shows the sequencing of parameter sets by period and session. M, T, L and S refer to the MANA, TR, large buyer and small buyer market types respectively. The subscripts index the parameter sets. The observation corresponding to session 4, period 4 is $[T \times L]_2$, which implies that the second parameter set is used to calibrate a TR market with large buyers. Other observations in Table 3.2 are analogously interpreted. To minimize ordering effects we alternated between TR and MANA markets.

Table 3.2. Distribution of Parameter Sets by Session and Period

		Session							
		1	2	3	4	5	6	7	8
	1	$[M\times L]_1$	$[M\times L]_1$	$[M\times L]_1$	$[M\times L]_1$	$[M\times S]_5$	$[M\times S]_5$	$[M\times S]_5$	$[M\times S]_5$
P	2	$[T\times L]_1$	$[T\times L]_1$	$[T\times L]_1$	$[T\times L]_1$	$[T\times S]_5$	$[T\times S]_5$	$[T\times S]_5$	$[T\times S]_5$
e	3	$[M\times L]_2$	$[M\times L]_2$	$[M\times L]_2$	$[M\times L]_2$	$[M\times S]_6$	$[M\times S]_6$	$[M\times S]_6$	$[M\times S]_6$
r	4	$[T\times L]_2$	$[T\times L]_2$	$[T\times L]_2$	$[T\times L]_2$	$[T\times S]_6$	$[T\times S]_6$	$[T\times S]_6$	$[T\times S]_6$
i	5	$[M\times L]_3$	$[M\times L]_3$	$[M\times L]_3$	$[M\times L]_3$	$[M\times S]_7$	$[M\times S]_7$	$[M\times S]_7$	$[M\times S]_7$
o	6	$[T\times L]_3$	$[T\times L]_3$	$[T\times L]_3$	$[T\times L]_3$	$[T\times S]_7$	$[T\times S]_7$	$[T\times S]_7$	$[T\times S]_7$
d	7	$[M\times L]_4$	$[M\times L]_4$	$[M\times L]_4$	$[M\times L]_4$	$[M\times S]_8$	$[M\times S]_8$	$[M\times S]_8$	$[M\times S]_8$
	8	$[T\times L]_4$	$[T\times L]_4$	$[T\times L]_4$	$[T\times L]_4$	$[T\times S]_8$	$[T\times S]_8$	$[T\times S]_8$	$[T\times S]_8$

Table 3.3 summarizes the profits π, prices p, number of trades n and efficiency EF predictions under the competitive equilibrium and monopsony solutions, indexed by C and MO superscripts respectively. S_A, S_B and B subscripts index the two seller types and the buyer type. Hence π_B^{MO} refers to the expected profits of a buyer at the monopsonist solution. Other column headers are analogously interpreted. The expected

profits in the L markets are scaled by the exchange rate between earnings in the L and S markets ($E\$18$ in L market = $E\$8$ in S market) to facilitate comparison. In the L market prices at the competitive equilibrium span an interval. p_C listed in Table 3.3 is the mid-point of this interval.

EF^{MO} refers to the expected market trading efficiency at the monopsony outcome. The expected trading efficiency at the competitive outcome is always one. Monopsony is not expected to induce a loss in trading efficiency EF in the L treatment. Price will deviate to $p^{MO} < p^C$, but trade volume will remain unchanged at $n^{MO} = n^C = 8$ and all sellers will trade. Monopsony will change the distribution of surplus in favor of buyers, but not at the cost of efficiency. In the S market monopsony will induce a fall in trading volume and a $4 - 18\%$ loss in trading efficiency. The difference between p^C and p^{MO} is greater in the L treatment. When buyers are able to exert market power their supercompetitive profits will be larger than in the S treatment.

Table 3.3. Model Predictions

Market	$\pi^C_{S_A}$	$\pi^{MO}_{S_A}$	$\pi^C_{S_B}$	$\pi^{MO}_{S_B}$	π^C_B	π^{MO}_B	p^C	p^{MO}	EF^{MO}	n^C	n^{MO}
$[M \times L]_1$	13.82	0.00	13.82	0.00	4.25	59.54	5.21	1.76	1.00	8	8
$[M \times L]_2$	13.96	0.00	17.44	0.00	4.59	67.39	7.04	3.56	1.00	8	8
$[M \times L]_3$	14.17	0.00	14.17	0.00	1.02	57.69	5.70	2.87	1.00	8	8
$[M \times L]_4$	18.43	0.00	15.80	0.00	10.22	78.69	5.30	2.67	1.00	8	8
$[M \times S]_5$	20.00	20.00	20.00	10.00	20.00	38.00	20.00	18.00	0.95	5–7	4
$[M \times S]_6$	20.00	12.00	20.00	20.00	20.00	34.00	12.00	10.00	0.94	5–6	4
$[M \times S]_7$	20.00	12.50	20.00	20.00	20.00	33.00	17.00	14.50	0.93	6–8	4
$[M \times S]_8$	20.00	20.00	20.00	12.50	20.00	29.00	19.00	17.50	0.82	7–8	4
$[T \times L]_1$	16.90	0.00	20.91	0.63	16.92	91.28	4.54	1.16	1.00	8	8
$[T \times L]_2$	18.40	1.96	20.56	0.00	4.59	78.59	7.04	2.93	1.00	8	8
$[T \times L]_3$	24.10	2.56	18.47	0.00	10.22	90.24	5.30	2.22	1.00	8	8
$[T \times L]_4$	20.07	2.67	14.50	0.00	3.82	67.62	5.57	2.67	1.00	8	8
$[T \times S]_5$	20.00	12.00	20.00	20.00	20.00	32.00	20.00	18.00	0.88	5–6	4
$[T \times S]_6$	20.00	20.00	20.00	11.00	20.00	34.00	12.00	9.00	0.86	5–6	4
$[T \times S]_7$	20.00	20.00	20.00	12.00	20.00	36.00	17.00	15.00	0.96	4–6	4
$[T \times S]_8$	20.00	14.00	20.00	20.00	20.00	28.00	19.00	18.00	0.89	5–7	4

3.1.4 Procedural Details

When the market opens in the first round a seller knows the cost and credits supplied by her three abatement projects. She uses this information to make a single offer, consisting of project name and price. A buyer knows the redemption values and credits associated with the six projects on offer. He uses this information to make multiple bids. After the market closes bids and offers are ranked by project. Bids are ranked in descending order (highest price first) and offers in ascending order. Bids and offers of equal rank and that have bid price exceeding offer price are paired by project. These are considered the successful bids and offers for the round. Unit bid and offer prices are calculated for these successes and then compiled into demand and supply schedules. The point or interval where schedules intersect determines the uniform market price p per credit. Let π_S, q_S and $C(\cdot)$ be the profit, credits supplied and the project cost for a representative seller S. Let π_B and $R(\cdot)$ be the profit and redemption value of a representative buyer B. Let B_S be the set of sellers who contract with B. Then $\sum_{S \in B_S} q_S$ is the quantity of credits bought by B. The uniform market price p is used to calculate profits for the round where

$$\pi_S = p \cdot q_S - C(q_S) \tag{3.1}$$

$$\pi_B = R\left(\sum_{S \in B_S} q_S\right) - p \cdot \sum_{S \in B_S} q_S \tag{3.2}$$

In the next round when the market opens the seller is informed of her offer status and the project profitability at current prices. The buyer is informed of of his bid(s) status and the redemption value earned from successful bids. Subjects are also informed of

the market price and all other bids and offers made in the previous round. While the market is open the buyer makes new bids wile the seller amends, resubmits or deletes her old offer or makes a new offer. When the market closes bids and offers are again ranked and a new market price determined. The process continues until the period ends. Termination criteria are either that eight rounds have passed or that the market price remains unchanged over two consecutive periods. Including instructions, sessions took about two hours to complete. Earnings were denominated in experimental dollars, which were converted into US dollars at E$18 to US$1 in the L market type and E$8 to US$1 in the S market type. Earnings ranged from US$10 to US$83 with a mean of about US$21. Experiment instructions are available on request.

3.2 Hypotheses

3.2.1 Under Size & Market Power Treatment

Let us first look at the implications of buyer size on outcome. The competitive equilibrium is, as always, most efficient and is the optimal regulator outcome. When buyers are large the competitive equilibrium is marked by the intersection of $d_L d'_L$ and ss' in Figure 3.1. The price lies in the interval $[p_L^L, p_L^H]$ and $q_L = q_{max}$ credits will be traded. The minimum buyers' surplus is $a + b$ and the maximum buyers' surplus is $a + b + c + d$. The minimum sellers' surplus in the L market is $e + f + g + h + i$ and the maximum sellers' surplus is $c + d + e + f + g + h + i$. When buyers are smaller then the competitive equilibrium price is p_S and credits traded lie in the interval $[q_S^L, q_S^H]$. The buyers' surplus is $b + c + e + f$ and the sellers' surplus is $h + i$.

Buyer profits are maximized when they collude perfectly. Under perfect collusion buyers act as a monopsonist maximizing aggregate profits through manipulation of both credit price and trade volume. In the L market type joint buyers' profits or surplus is maximized by restricting price to p_S and credit purchase to q_L. Buyers' surplus is $a+b+c+d+e+f+g$ and sellers' surplus is h_i. When buyers are small joint profits are maximized when some $q < q_S^L$ credits are traded at price p. Buyers' surplus is $b+c+e+h$ and sellers' surplus is i.

We expect the monopsony / collusive solution and competitive equilibrium to be the extreme outcomes in the market. It is possible for prices to rise above the competitive equilibrium price, but that implies market power lying with the many sellers. Such an outcome requires much coordination among sellers, which though feasible is unlikely.

When the buyers do not collude the outcomes will lie between the monopsonist and competitive extremes. Competition over credits will raise prices above monopsony levels but not to the level at the competitive equilibrium. When buyers are large we expect prices to lie between p_S and p_L^H and when buyers are small, prices will lie between p and p_S. The big question is whether prices will be closer to competitive levels in the L or S markets. In other words we are interested in whether the large or small buyers are able exert more market power. On one hand, for large buyers the relative scarcity of credits is greater, which lays the groundwork for greater competition for resources. When buyers compete the only way for a buyer to garner more resources is to outbid the other. Since large buyers need a greater credit stock $\sum_{S \in B} q_S$ they may bid prices up to a greater extent than the small buyers. High prices will result in a smaller share of the gains from trade going to the buyers, which implies that small buyers will exercise more market

power than large buyers. On the other hand, supercompetitive profits from enforcing market power are higher for large buyers. If they realize this potential for greater payoffs large buyers will have greater incentive to coordinate their bidding strategies and hence will exert more market power than small buyers. Since subjects were not allowed to communicate coordination is tacit; in the form of responses to observed behavior.

Traded quantities in the L and S markets will be q_L and $q < q_S^L$ respectively. Market power will cause a reduction in credits traded when buyers are small, but not when they are large. Under the small buyer market quantity and price distortion are expected, but in the large buyer market only price distortion is expected. As a result it is expected that the large buyer market will be more efficient than the small buyer market. We propose four related testable hypotheses on the relationship between size and market power.

Hypothesis 1. *Credit prices will be closer to the competitive equilibrium price in the large buyer treatment*

Hypothesis 2. *Small buyers will capture a greater proportion of the gains from trade and hence exercise more market power*

Hypothesis 3. *Quantity of trade will be lower than at the competitive equilibrium in the small buyer treatment. There will be no quantity distortion in the large buyer environment*

Hypothesis 4. *The market will be more efficient when buyers are large*

Validation of Hypotheses 1 and 2 will support the thesis that the large buyers competed instead of cooperating when procuring the limited credit stock. Non-validation

of these hypotheses supports the opposite thesis that the large buyers cooperate in the procurement of the scarce resource in recognition of the greater payoffs from doing so.

3.2.2 Under the Market Institution Treatment

We can think of the cost-effectiveness of a project as its cost per credit supplied, $C(q_S)/q_S$. When $C(q_S)/q_S$ is lower the project's cost-effectiveness is higher. Dividing RHS and LHS of (3.1) by q_S it is apparent that the most cost-effective projects are also the most profitable. Sellers will prefer to offer their most cost-effective projects to the market. In many cases, though not always, the most cost-effective projects will also supply the maximum credits.

Buyers purchase credits from multiple sellers until they own the profit maximizing quantity $q^*_B = \sum_{S \in B^*_S} q_S$ where B^*_S is a project portfolio supplying q^*_B. The greater the number of trades, the greater the uncertainty over whether the buyer will manage to buy q^*_B credits in a round. The buyer's preference will be for fewer trades, which implies more bids for high credit projects.

The MANA and TR markets differ in the manner in which projects are appraised and credits allocated. Since $q^M_S = E - V$ and $q^T_S = E/t$ more credits are allocated to projects with high mean and low variability in the MANA market and V does not affect credit generation in the TR market. The regulator, seeking to control credit risk, prefers projects with high q^M_S, which may be considered high quality projects. Cost-effectiveness from the regulator's point of view is $CE_S = C(q^M_S)/q^M_S$. We have four hypotheses on the types of projects traded in the two market institutions.

HYPOTHESIS 5. *Projects traded in the MANA market will be of a higher quality than projects traded in the TR market*

HYPOTHESIS 6. *The aggregate variability of credits traded in the MANA market will be lower*

HYPOTHESIS 7. *The aggregate expected value of credits traded in the TR market will be higher*

HYPOTHESIS 8. *Projects traded in the MANA market will be more cost-effective from the regulator's point of view*

3.3 Tests

3.3.1 For Market Efficiency

A common way to assess a market's trading efficiency is to compare gains from trade in the experimental market to gains from trade at the competitive equilibrium. The proportional difference between the two is typically used as the measure of market efficiency by environmental experimentalists (see Cochard et al. 2005; Muller et al. 2002; Spraggon 2002). Gains from trade is the surplus that accrues from trade and is defined as the difference between aggregate profit after trade and aggregate profit at autarky. If π^{AU} is profit under autarky then market efficiency EF is defined as

$$EF = \frac{\sum \pi - \sum \pi^{AU}}{\sum \pi^{C} - \sum \pi^{AU}} \tag{3.3}$$

$EF = 0$ when there are no gains from trade and $EF = 1$ when the market outcome is competitive and all gains from trade are realized.

The *raison d'etre* of the MANA and TR markets is the least cost attainment of the regulator's environmental target. Under stochasticity, Ghosh and Shortle (2008) show that optimality requires the regulator's cognizance of a trade-off between the mean E and variance V of abatement. In this experiment the trade-off is recognized through the MANA market credit definition $q_S^M = E - V$. This definition may be considered a quality index because it scores over the two abatement attributes that the regulator must control. The more credits q_S^M generated by a project the higher its quality. If $\tilde{q}_S^M$ is the credits sold by seller S at the regulator's optimum then $\sum \tilde{q}_S^M$ is the optimal level of nonpoint source abatement. The market may be judged by its success in achieving this optimal level of trade. We define the market performance measure

$$POA = \frac{\sum q_S^M}{\sum \tilde{q}_S^M} \tag{3.4}$$

where POA is the Proportion of Optimum Abatement. POA measures the market's ability to meet the regulator's loadings target and may be considered a measure of physical efficiency. When $POA = 0$ the market is unable to generate any abatement and when $POA = 1$ it generates the optimal level. When $POA > 1$ the market causes over-abatement. The market's physical efficiency increases as $POA \to 1$. POA is based on $P - MAR$, an efficiency measure proposed by Cason and Gangadharan (2005) to compare the efficiency of auctions at reducing nonpoint source pollution.

3.3.2 For Market Power

Market power may be defined as the ability to profitably deviate prices from the competitive equilibrium (Mas-Colell et al. 1995, pg. 383). It can be measured by comparing realized prices and profits to competitive equilibrium levels. Isaac et al. (1984) devised the Index of Monopoly Trading Effectiveness, a unitless measure of market power based on seller profits. We devise an analogous Buyers' Index of Monopsony Trading Effectiveness based on buyer profits, I^B:

$$I^B = \frac{\sum_B \pi_B - \sum_B \pi_B^C}{\sum_B \pi_B^{MO} - \sum_B \pi_B^C} \tag{3.5}$$

$\sum_B \pi_B$ in (3.5) is aggregate buyer profit. $I^B = 0$ when buyers earn profits equal to the competitive equilibrium level and $I^B = 1$ when they earn profits equal to the monopsony level. When $0 < I^B < 1$ we infer that buyers collude to exert market power, which increases as $I^B \to 1$. I^B may be negative either when price exceed the competitive price level, or when trading volumes fall below competitive levels but prices remain near p^C. Prices exceed the competitive level when sellers exert market power. The second scenario of low trade volumes but competitive prices might imply trading inefficiency. We also propose a Sellers' Index of Monopsony Trading Effectiveness based on seller profits, I^S:

$$I^S = \frac{\sum_S \pi_S - \sum_S \pi_S^{MO}}{\sum_S \pi_S^C - \sum_S \pi_S^{MO}} \tag{3.6}$$

where $\sum_S \pi_S$ is aggregate seller profit. $I^S = 0$ when when buyers collude perfectly and generate the monopsony outcome. $I^S = 1$ when seller earnings equal the competitive

equilibrium level. The sellers exert market power when $I^S > 1$ and buyers exert market power when $0 < I^S < 1$, increasing as $I^S \to 0$.

Drawing directly from the definition in Mas-Colell et al. (1995), another indication of market power is the credit price. Buyers will try to ensure that $p \to p^{MO}$ because such price movement is profitable. The degree by which p moves towards p^{MO} provides a measure of market power. We propose the Price Index of Monopsony Trading Effectiveness based on credit price, I^p:

$$I^p = \frac{p - p^{MO}}{p^C - p^{MO}} \tag{3.7}$$

When buyers collude perfectly and push prices to the monopsony level then $I^p = 0$. On the other hand when sellers are able to resist the downward pressure on price and generate the competitive outcome then $I^p = 1$. Buyers exert market power when $0 < I^p < 1$, increasing as $I^p \to 0$. Sellers exert market power when $I^p > 1$.

3.3.3 For Abatement Quality

A final set of tests evaluate the quality of projects traded in the market. Four related measures supply information on the project quality as measured by their physical attributes. These are the total abatement credits traded in the market, $\sum_S q_S$, the expected abatement from projects traded in the market, $\sum_S E_S$, the total variability of traded projects, $\sum_S V_S$ and the mean coefficient of variation of traded projects,

$\overline{CV} = \sum_S E_S/N_S\sqrt{V_S}$, where N_S is the number of traded projects. The final test measures the aggregate regulator-defined cost effectiveness of traded projects, $\sum_S CE_S = \sum_S C(q_S^M)/q_S^M$.

3.4 Results and Discussion

In the subsequent discussion metrics measuring performance in the MANA, TR, large buyer and small buyer markets are subscripted by M, T, L and S respectively.

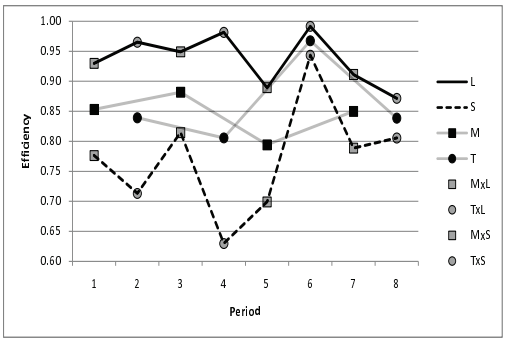
3.4.1 Efficiency

Trading efficiency EF across periods and treatments is shown in Figure 3.2(a) and ranges from 0.63 in the $[T \times S]_6$ market type to 0.99 in the $[T \times L]_7$ market type. Mean efficiency in the L and S markets are $\overline{EF}_L = 0.94$ and $\overline{EF}_S = 0.77$ respectively. According to the nonparametric Wilcoxon signed rank test,[4] $\overline{EF}_L$ is significantly greater than $\overline{EF}_S$ at the 1% level. The Wilcoxon test also indicates that $\overline{EF}_L < 1$ and $\overline{EF}_S < 1$ at the 1% level which implies that market outcomes in the L and S markets are significantly less efficient than the competitive equilibrium outcome. *The Wilcoxon results lend strong support to Hypothesis 4, which states that the market will be more efficient under the L treatment.*

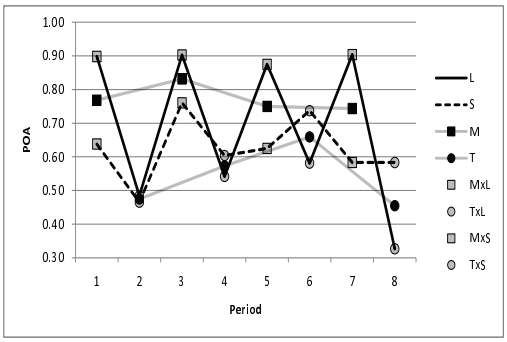
Mean market efficiency in the M and T markets are $\overline{EF}_M = 0.84$ and $\overline{EF}_T = 0.86$ respectively. Wilcoxon test results indicate that the difference between the $\overline{EF}_M$ and

[4] An advantage to the Wilcoxon test is that it does not require assumptions on the distribution of the data. It does, however, require that data points are independent and that the two samples are drawn from identical populations (Hays and Winkler 1971).

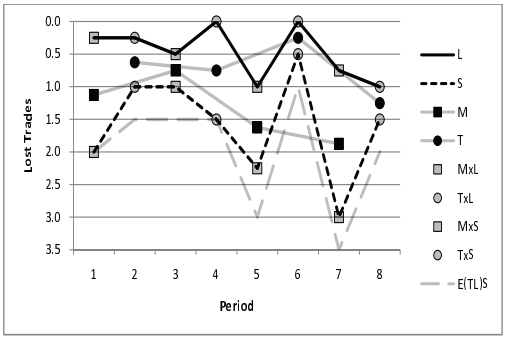
Fig. 3.2. Efficiency Indices



(a) Market Efficiency, EF



(b) Proportion of Optimal Abatement, POA



(c) Transaction Efficiency Loss, TL

$\overline{EF}_T$ is not significant at the 10% level. Efficiency in both markets is less than at the competitive equilibrium with $\overline{EF}_M < 1$ and $\overline{EF}_T < 1$ at the 1% level.

A shortcoming of the Wilcoxon and other univariate tests is that they supply limited information on factors affecting the variable of interest. We estimate a series of random effects panel data models to quantify the relationships between test results and treatments. The results are presented in Table 3.4. The estimates are corrected for serial correlation and heteroscedasticity through application of the Newey and West (1987) procedure. The independent variables in all models are dummies for treatment effects and a time trend specified as 1/Period Number. Dum_{Inst} corresponds to the market institution treatment, equalling 1 under the T treatment and 0 otherwise. Dum_{Size}

corresponds to the buyer size treatment, equalling 1 under the L treatment and 0 otherwise. As noted by Cason and Gangadharan (2003) the 1/Period time trend specification allows the intercept to be interpreted as asymptotic market efficiency. The dependent

Table 3.4. Panel Regression Models for Efficiency Tests

Dependent Variable	EF in Period (Model 1)	POA in Period (Model 2)	TL in Period (Model 3)
Intercept	0.76^{*} (0.04)	0.75^{*} (0.04)	2.16^{*} (0.32)
Dum_{Inst}	0.02 (0.03)	-0.24^{*} (0.03)	-0.72^{*} (0.23)
Dum_{Size}	0.16^{*} (0.03)	$0.06^{\dagger}$ (0.03)	-1.12^{*} (0.25)
1/Period	0.00 (0.07)	-0.02 (0.07)	-0.61 (0.42)
R-Squared	0.30	0.38	0.36
Number of Obs.	64	64	64

Standard errors are in parentheses. * and † indicate that coefficients are significantly different from zero at the 1% and 10% levels. The session acts as the random effect, corrected for through the Newey and West (1987) procedure.

variable in Model 1 is EF in every period, which implies (8 periods$\times$ 8 sessions)=64 observations. The coefficient on Dum_{Size} is significant, corroborating the Wilcoxon test result that the market is significantly more efficient when buyers are large. EF is slightly higher in the T treatment than the M, but the effect is small and insignificant. The time trend has no effect on EF. An increase in subject experience does not increase market efficiency. We discuss this point in greater detail in Section 3.4.4.

The regulator wants the market to generate the optimal level of abatement through trade in $\sum \tilde{q}_S^M$ credits. POA measures the market's ability to create this optimal level of trade and abatement and is a measure of the market's physical efficiency. Figure 3.2(b) shows POA across the different periods and treatments. Abatement in the M market is closer to the regulator's optimum than abatement in the T market with $\overline{POA}_M = 0.77$ and $\overline{POA}_T = 0.54$. The Wilcoxon test indicates that the difference in physical efficiency

between the M and T markets is significant at the 1% level. *Aggregate abatement quality is higher in the M market than in the T market, which supports Hypothesis 5.*

Similar quantities of aggregate abatement are generated in the L and S markets with $\overline{POA}_L = 0.69$ and $\overline{POA}_S = 0.62$. The difference is not significant at the 10% level according to the Wilcoxon test. This result belies the expectation under Hypothesis 3 that $\overline{POA}_L > \overline{POA}_S$. The other two expectations under Hypothesis 3 are that $\overline{POA}_L = 1$ since the quantity of credits traded is not expected to deviate from the (regulator-optimal) competitive equilibrium level, and $\overline{POA}_S < 1$. Wilcoxon tests show that $\overline{POA}_L < 1$ and $\overline{POA}_S < 1$ at the 1% level of significance. *Overall The evidence supporting Hypothesis 3 is weak.*

The Wilcoxon test result on the market institution treatment effect on POA is corroborated by Model 2 in Table 3.4. $Dum_{Inst} = -0.24$ and significant at 1%, implying that POA under the T treatment is 24% less than under the M treatment. However the Wilcoxon result on the (non-)impact of buyer size is not corroborated. Dum_{Size} is positive and significant at the 10% level, which weakly supports the expectation in Hypothesis 3 that the physical efficiency of the market is higher when buyers are large. The time trend coefficient is negative but insignificant, implying that a market's physical efficiency does not increase as subjects gain experience.

Transaction volume or the number of trades n is another indicator of market efficiency. Experimental markets are transactionally efficient when $n = n^C$, the number of trades at the competitive equilibrium. Two factors affect transaction volume in this experiment. The first is monopsony and the second consists of unobservables that hinder the contracting process. Expected transaction volumes under each market type at the

competitive equilibrium and under monopsony are listed in Table 3.3. We define the transaction efficiency loss TL of a market as $TL = n^C - n$. $TL = 0$ when the market is transactionally efficient and increases as inefficiency increases.

The L market is calibrated such that it will not lose transaction efficiency because of monopsony: $E(TL)_L = 0$. Realized TL_L per period is in the 0-1 range as seen in Figure 3.2(c) with mean, $\overline{TL}_L = 0.47$, significantly different from zero at 1%. The expected loss in transaction efficiency due to monopsony under the S treatment is 1-3.5 units per period as marked by the $E(TL)_S$ line in Figure 3.2(c). Realized TL_S ranges from $0.5 - 3.0$ units with mean $\overline{TL}_S = 1.59$, which is lower than the monopsony prediction. Transaction volume n is significantly higher in the L treatment with $\overline{TL}_L > \overline{TL}_S$ at the 1% level. Significance levels were determined through Wilcoxon tests. These results validate the expectations in Hypothesis 3 that trade levels are higher in the L treatment and that trade volume in the S treatment is less than the competitive equilibrium volume. The results do not validate the third expectation in Hypothesis 3 that trade volumes in the L market are at the competitive equilibrium level.

There is a significant difference (at the 5% level) in n between the two market institutions. Transaction volume is significantly higher in the T market with $\overline{TL}_T = 0.72$ and $\overline{TL}_M = 1.34$. The difference might be explained by looking at the $E(TL)_S$ line in Figure 3.2(c). Note that $E(TL)_S$ is lower in the odd-numbered periods, which correspond to the $M \times S$ market type, which implies that $E(TL)_{M\times S} > E(TL)_{T\times S}$ under monopsony. Even though the expected trade volume in $M \times L$ and $T \times L$ markets is identical, lower expected trade volume in $M \times S$ markets might be sufficient to ensure that $\overline{TL}_M > \overline{TL}_T$.

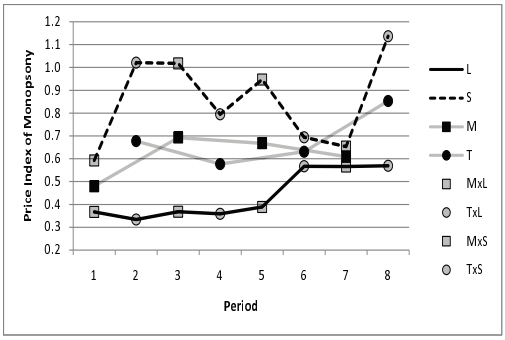
In Model 3 in Table 3.4 we regress the TL statistic against the market institution and buyer size treatments. Both treatments are found to significantly affect trading volume at the 1% level in the manner indicated by the Wilcoxon tests described above. Trading volume increases over periods, but this effect is not significant.

3.4.2 Market Power

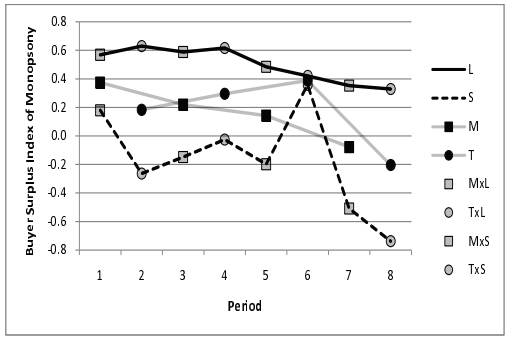
The first measure of market power analyzes is the price index of monopsony trading effectiveness I^p, defined in (3.7). When $I^p = 1$ then prices are perfectly competitive and the degree of monopsony increases as $I^p \to 0$. Mean I^p values by market type are shown in Figure 3.3(a). The price index for the L market I^p_L is lower than the price index for the S market I^p_S. In the mean $\bar{I}^p_L = 0.44$ and $\bar{I}^p_S = 0.86$. With a p-value of 0.79 the hypothesis that $\bar{I}^p_S = 1$ cannot be rejected. Also $\bar{I}^p_L < \bar{I}^p_S$ at the 1% level according to the Wilcoxon test. The test results allow the inference that credit prices in the L treatment are significantly lower than credit prices in the S treatment, which are not significantly different from the competitive equilibrium price. *The results run contrary to the expectation in Hypothesis 1 that credit prices will be higher in the L treatment.* The non-validation of Hypothesis 1 implies that large buyers cooperated instead of competing, exerted market power, drove prices lower than small buyers could and thereby increased their profits. The large buyers realized that profits through cooperation were higher than profits under competition and acted accordingly.

The price indices differ slightly across the two market institutions with $I^p_M = 0.61$ and $I^p_T = 0.68$. The difference is not significant at the 1% level. The prices are significantly below competitive equilibrium levels at the 1% level.

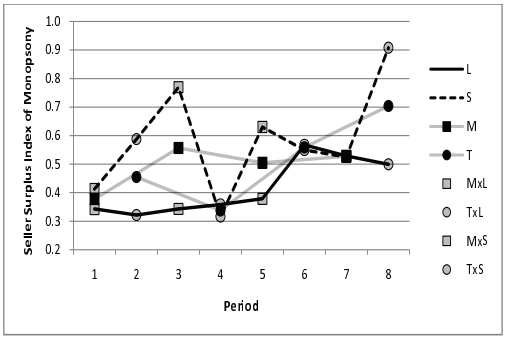
Fig. 3.3. Indices of Monopsony Trading Effectiveness



(a) Price-Based Index, I^p



(b) Buyers' Surplus Index, I^B



(c) Sellers' Surplus Index, I^S

We run regressions of the three market power statistics on treatment effects and present the results in Table 3.5. The regression results for I^p are presented in Model 4. The coefficient on Dum_{Size} is negative and significant at 1%, which corroborates the Wilcoxon test results that market price is significantly lower under the L treatment and provides further evidence against Hypothesis 1. The coefficient on the time trend is negative, indicating a decrease in buyers' market power as the experiment progresses and subjects gain experience, but this impact is not significant. The coefficient on Dum_{Inst} is not significant.

The second measure of monopsony is the Buyers' Index of Monopsony Trading Effectiveness I^B. When aggregate buyers' surplus reaches competitive equilibrium levels

Table 3.5. Panel Regression Models for Monopsony Statistics Tests

Dependent Variable	I^p in Period (Model 4)	I^B in Period (Model 5)	I^S in Period (Model 6)
Intercept	0.91^* (0.1)	$-0.34^\dagger$ (0.18)	0.67^* (0.11)
Dum_{Inst}	0.04 (0.07)	0.06 (0.11)	-0.01 (0.07)
Dum_{Size}	-0.42^* (0.09)	0.67^* (0.12)	$-0.17^\dagger$ (0.09)
1/Period	-0.21 (0.16)	$0.40^\dagger$ (0.22)	-0.23 (0.15)
R-Squared	0.32	0.38	0.09
Number of Obs.	64	64	64

Standard errors are in parentheses. * and † indicate that coefficients are significantly different from zero at the 1% and 10% levels. The session acts as the random effect, corrected for through the Newey and West (1987) procedure.

then $I^B \to 0$ and when surplus reaches monopsony levels then $I^B \to 1$. Figure 3.3(b) indicates that buyers' surplus is higher under the L treatment than under the S treatment. In the mean $\bar{I}^B_L = 0.50$ and $\bar{I}^B_S = -0.17$ and the difference between the two is significant at the 1% level according to the Wilcoxon test. With a p-value of 0.13, the evidence against the assertion that $\bar{I}^B_S = 0$ is weak, which implies that the buyers' surplus equals the perfectly competitive level under the S treatment. In contrast, $\bar{I}^B_L > 0$ at the 1% level, indicating significant exercise of market power by buyers in the L treatment. *These results do not validate Hypothesis 2, which states that small buyers will exert more market power to capture a greater proportion of the gains from trade.* Instead the results support the alternate thesis that large buyers respond to the greater financial rewards from cooperation by suppressing prices more than the small buyers.

The proportion of surplus captured by buyers in the M and T markets is very similar with $\bar{I}^B_M \approx \bar{I}^B_T \approx 0.17$. The Wilcoxon test provides no evidence that $\bar{I}^B_M$ and $\bar{I}^B_T$ differ significantly. We do not reject the hypothesis that $\bar{I}^B_T > 0$ at the 10% level and that $\bar{I}^B_M > 0$ at the 5% level. Surplus capture by buyers in the M and T markets is greater

than competitive equilibrium levels but the significance of the result is marginal. The results of the regression of I^B on treatments and the time trend are given in Table 3.5. The buyer size dummy is positive and significant at 1% invalidating Hypothesis 2 and supporting the Wilcoxon test result that greater market power is exerted by large buyers. The time trend is significant at 1%, which implies a reduction in I^B and a move towards the competitive equilibrium outcome as subjects gain experience. Sellers are become better at resisting the downward pressure on price exerted by buyers and thereby reduce buyers' surplus.

The final measure of market power considered is the Sellers' Surplus based Index of Monopoly Effectiveness (3.6). At the monopsony outcome $I^S = 0$ and at the competitive equilibrium outcome $I^S = 1$. Mean values of I^S across treatments and periods is shown in Figure 3.3(c). Surplus capture by sellers is significantly higher at the 5% level under the S treatment ($\bar{I}^S_S = 0.59$) than in the L market ($\bar{I}^S_L = 0.42$), once again invalidating Hypothesis 2.

Surplus capture by sellers is similar under the two market institution treatments with $\bar{I}^S_M = 0.49$ and $\bar{I}^S_T = 0.51$. The difference between the two statistics is not significantly different from zero at the 10% level. Mean outcomes under all treatments differ significantly from both the competitive equilibrium and monopsony outcomes.

The results from the regression of I^S on the experimental treatments and time trend are provided in Model 6 of Table 3.5. Similar to the results on the other monopsony measures Dum_{Size} is a significant regressor (though only at the 10% level). Sellers' surplus capture is significantly less when buyers are large. The coefficient on the time trend

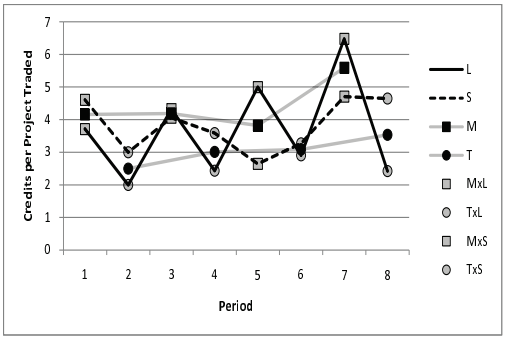
is negative, implying an increase in buyers' market power as the experiment progresses, but the impact is not significant.

3.4.3 Abatement Quality

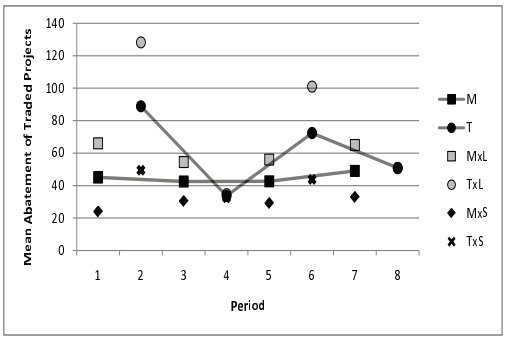
The quality and quantity of abatement generated by the market indicates its success in meeting the regulator's environmental target. The proportion of optimal abatement POA captured the market's ability to reach this target. In the following analysis qualitative differences in the abatements achieved by the M, T, S and L market types are explored. Consider the quality of the abatement projects traded in the different market types. Project quality is measured by the number of credits allocated to the project in the $MANA$ market, q^M . Figure 3.4(a) shows the differences in project quality across the four market types. The difference in mean credits $\bar{q}^M$ per projects traded seems significant when comparing the M and T markets, but not when comparing the L and S markets. Mean abatement credits supplied per project in the M market is $\bar{q}^M_M = 4.44$ and in the T market is $\bar{q}^M_T = 3.03$. Mean abatement credits per project in the L market is $\bar{q}^M_L = 3.66$ and in the S market is $\bar{q}^M_S = 3.82$. The Wilcoxon test indicates that $\bar{q}^M_M > \bar{q}^M_T$ at the 1% level. There is no evidence supporting the assertion that $\bar{q}^M_L \neq \bar{q}^M_S$.

Model 7 in Table 3.6 reports results of the regression of $\bar{q}^M$ on treatment effects. The market institution dummy Dum_{Inst} coefficient is negative and significant at 1%, which supports the Wilcoxon test result that $\bar{q}^M_L$ is significantly less than $\bar{q}^M_S$. *The results lend support to Hypothesis 5, which states that project traded in the $MANA$ market are of a higher quality than projects traded in the TR market.* The two characteristics

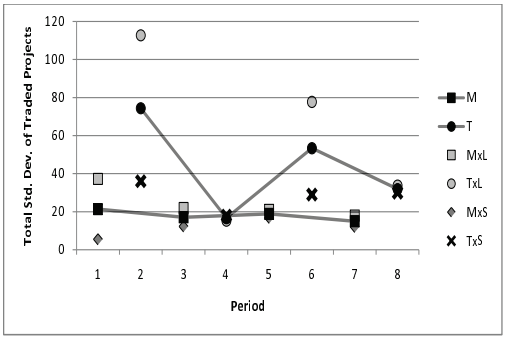
Fig. 3.4. Quality and Cost Effectiveness of Traded Projects



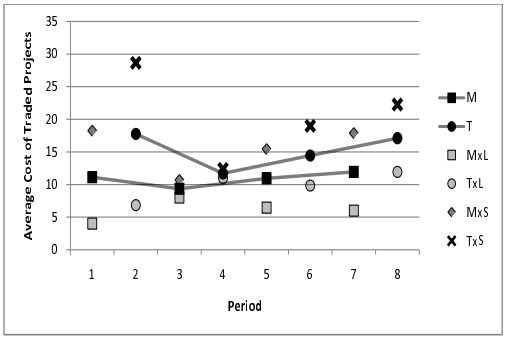
(a) Average Credits per Trade, $\bar{q}^M$



(b) Mean Abatement from Trade, $\sum E$



(c) Standard Deviation of Traded Abatement, $\sum SD$



(d) Unit Cost of Traded Abatements, $\sum C(q^M)/q^M$

of nonpoint abatement that are of interest to the regulator are its expectation and variance. The regulator prefers projects with high $q^M = E - V$ and does not have preferences over E independent of V. Under TR market rules E is the only project attribute of importance. Credits awarded to a project in the T market increase as E increases, but are independent of V. Given that buyers prefer projects that supply more credits (because this implies fewer trades are necessary to reach their optimal portfolio) projects with high E will have maximum demand in the T market. These projects also tend to be cost-effective in the T market. There is no such preference for high E projects independent of V in the M market. As stated in Hypothesis 7 the aggregate expected

Table 3.6. Panel Regression Models for Abatement Quality Statistics

Dependent Variable	Q in Period (Model 7)	E in Period (Model 8)	SD in Period (Model 9)	AC in Period (Model 10)
Intercept	4.92^{*} (0.34)	23.14^{*} (4.63)	-1.37 (4.92)	15.54^{*} (0.99)
Dum_{Inst}	-1.56^{*} (0.29)	18.63^{*} (5.48)	29.12^{*} (5.79)	4.53^{*} (1.00)
Dum_{Size}	-0.16 (0.20)	32.84^{*} (5.05)	22.28^{*} (4.79)	-10.06^{*} (0.93)
1/Period	-0.95^{**} (0.37)	$12.46^{\dagger}$ (6.91)	19.86^{*} (7.00)	0.79 (1.62)
R-Squared	0.41	0.46	0.45	0.71
Number of Obs.	64	64	64	64

Standard errors are in parentheses. *, ** and † indicate that coefficients are significantly different from zero at the 1%, 5% and 10% levels. The session acts as the random effect, corrected for through the Newey and West (1987) procedure.

abatement from traded credits in the T market is expected to be significantly higher than in the M market.

Let the aggregate expected value of traded credits in a market be $\mu = \sum E$. In the M market the aggregate expected value is $\mu_M = 44.78$ and in the T market it is $\mu_T = 61.43$. The Wilcoxon test indicates that the difference between the two means is significant at the 1% level, which supports Hypothesis 7. In the corresponding regression model (see Model 8 in Table 3.6) the market institution dummy is significant at 1% lending further support to Hypothesis 7. *There is strong evidence that the aggregate expected abatement from traded credits is significantly higher in the T market.*

The difference between the aggregate expected value of traded credits in the L and S markets is also significant according to the Wilcoxon test and Model 8. However this result does not have inferential value because it arises from underlying parameter differences. In the L markets the range of expected abatement supplied by a project is $4 \leq E_L \leq 18$, while in the S market the range is $4 \leq E_S \leq 12$. There is no parameterization effect when comparing expected abatement in the M and T markets because parameters in $M - T$ market pairs are identical.

Ceteris paribus projects that supply low variance abatements are awarded more credits than projects with high V abatement in the M market. In the T market by contrast, credit creation is independent of the variance of project abatement. Since buyers demand high credit projects and sellers tend to find them cost-effective, we expect that aggregate projects traded in the M market will have lower V than projects traded in the T market, as stated in Hypothesis 6.

Let $\sigma^2 = \sum V$ be the total variance of all projects traded in a market. Mean total standard deviation in the M and T markets are $\bar{\sigma^2}_M = 18.09$ and $\bar{\sigma^2}_T = 44.06$ respectively and the difference between the two means is significant at the 1% level. The results of the corresponding regression of σ^2 on treatments are given in Model 9 in Table 3.6. The market institution dummy is significant at 1%, indicating that σ^2 is indeed higher in the T market. *The Wilcoxon test and the regression both support Hypothesis 6, implying that aggregate variance of nonpoint abatement is lower in projects traded in the M market.* The difference in $\tilde{\sigma}^2$ across the L and S markets is also significant at 1% according to both the Model 9 and the Wilcoxon test, but was the case with $\tilde{\mu}$, the difference stems from parameter differences and has no inferential value.

The final metric of abatement quality is the average cost of projects traded per credit supplied. The credits are defined according to the regulator's preference. Hence average cost is $AC = \sum C(q^M) / \sum q^M$. The market that supplies projects with lower average costs may be dubbed as cost-efficient. The *raison d'etre* of an emissions trading market is its perceived cost efficiency. Hence markets with higher cost efficiency may be considered as the 'better' market. As stated in Hypothesis 8 we expect the M market to be more cost efficient.

Mean average cost in the M market is $\overline{AC}_M = 10.84$ and average cost in the T market is $\overline{AC}_T = 15.25$. The difference between $\overline{AC}_M$ and $\overline{AC}_T$ is significant at the 1% level. Model 10 in Table 3.6 is contains the results of the regression of AC on the experimental treatments. *In the model the market institution dummy is significant and positive at 1%, which supports the Wilcoxon test result that average cost of credit generation is higher in the T market than in the M market. Both results provide strong support for Hypothesis 8.* As with the results on aggregate mean and standard deviation, average cost per credit differs significantly between the S and L treatments with $\overline{AC}_L = 8.02$ and $\overline{AC}_S = 18.08$. The difference is once again ascribed to parameterization.

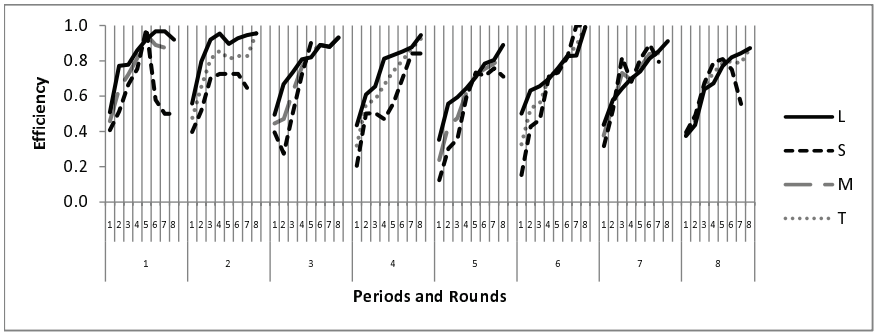
3.4.4 Efficiency Dynamics

So far we have focused on final market outcomes: the results in Sections 3.4.1–3.4.3 illuminate the state of the market after the period ended and all trades were executed. We now look at market evolution from round to round within a period. We identify a test statistic for round i by appending subscript i to the test statistic. For example, TS_4 refers to test statistic TS in round 4.

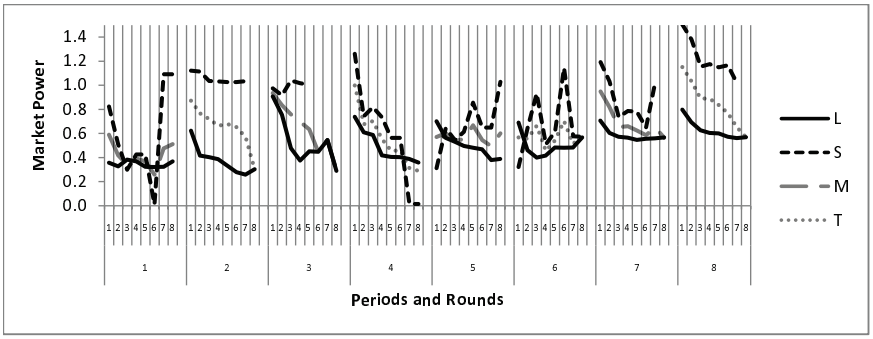
Market efficiency is measured through the trading efficiency statistic EF, as defined in (3.3). Mean trading efficiency across all rounds and periods is $\overline{EF} = 0.68$ and trends upwards as rounds increase from a first round low, $\overline{EF}_1 = 0.39$, to a final round high, $\overline{EF}_8 = 0.90$. Market efficiency increases over the course of the period as buyers and sellers learn more about the market, calibrate their bids and offers and organize successful trades. The evolution of EF across periods and rounds is shown in Figure 3.5(a).

There is a clear pattern where in each period $\overline{EF}$ increases by round. However, for a given round, there is no clear upward trend in $\overline{EF}$ as periods increase.

Fig. 3.5. Evolution of Market Efficiency and Market Power



(a) Efficiency per Round, EF



(b) Price-Based Monopsony Index per Round, I^p

A panel regression model, where EF per round is the dependent variable regressed against treatment effects and time trends for round and period, is more informative than Figure 3.5(a). The results for the regression are given in Model 11 in Table 3.7. The time trends are inverses of the Period and Round, which implies that the intercept can be interpreted as the asymptotic value of EF. Both the period and round coefficients are significant at the 5% level. The 1/Round coefficient is negative, indicating that market efficiency increases as rounds progress and confirming the visual evidence in Figure 3.5(a).

The 1/Period coefficient is positive, which implies that (holding 1/Round constant) EF decreases across periods. EF increases within periods and decreases across periods. In contrast, note that the 1/Period coefficient in Model 1 in Table 3.4 equals 0.00, implying that EF is not affected by period. The discrepancy is explained by the observation that EF in a period masks information on the round when the market closed. The EF-in-period data point corresponds to the EF_3 data point when the market closes in round 3 and the EF_8 data point when the market closes in round 8.

The institution treatment did not affect EF, but the Dum_{Size} coefficient was positive and significant at 1%, implying that EF was significantly higher in the L market. From Model 1 in Table 3.4 recall that EF-in-period was also higher in the L treatment. This was ascribed to the number of trades n being significantly closer to the competitive equilibrium level in the L market [see Model 3 in Table 3.4]. The same explanation holds for higher EF_i in the L market. The number of observations in Model 11 was 426. If each market ran for all eight rounds, given eight sessions of eight periods, we would have 64 observations in each round and 512 observations in all. Instead, five markets closed before after round three, 11 closed without needing rounds five onwards and 29 closed without needing round eight.

Market power in every round is measured by the Price Index of Monopsony Effectiveness I^p. There is a downward trend in I_i^p as the number of rounds increase, as examination of Figure 3.5(b) shows. Mean market efficiency across all periods is $\bar{I}^p = 0.64$, trending downwards as rounds increase from a first round high, $\bar{I}_1^p = 0.85$ to an eighth round low $\bar{I}_8^p = 0.49$. In the mean I_i^p decreases as rounds increase, implying an increase in market power exerted by buyers.

Table 3.7. Panel Regression Models of Within-Period Dynamics

Dependent Variable	EF in Round (Model 11)	I^p in Round (Model 12)
Intercept	0.73^{*} (0.05)	0.80^{*} (0.07)
Dum_{Inst}	0.03 (0.04)	0.04 (0.05)
Dum_{Size}	0.13^{*} (0.04)	-0.33^{*} (0.05)
1/Period	0.15^{**} (0.07)	-0.28^{*} (0.08)
1/Round	-0.52^{*} (0.04)	0.30^{*} (0.07)
R-Squared	0.47	0.24
Number of Obs.	426	421

Standard errors are in parentheses. * and ** indicate that coefficients are significantly different from zero at the 1% and 5% levels. The session acts as the random effect, corrected for through the Newey and West (1987) procedure.

Model 12 in Table 3.7 contains the results of a regression of I_i^p on treatment effects and time trends. As was the case with EF_i the institution treatment did not significantly affect I_i^p. The size treatment effect was significant at 1% and implied significantly greater market power exerted by large buyers. The result supports results in Model 4 in Table 3.5. Both time trend coefficients were also significant. I_i^p increases in Period and and decreases in Round, which implies that buyers' market power increases within periods but decreases across periods. As sellers gain experience they are able to better coordinate their offer strategies such that they resist buyers' market power. At the same time, their bargaining position gradually erodes as the length of time that a market remains open (measured by the number of rounds) increases. It is likely that some sellers are more susceptible to price pressure exerted by buyers and over time they crack, driving prices and the aggregate sellers' surplus down. There were five completely inefficient markets in Round 1 (with $EF_1 = 0$) when not a single bid-offer combination was successful. Market power statistics could not be calculated for these markets, which explains why there are 421 observations in Model 1 and 426 in Model 11.

3.5 Conclusion

We have two sets of results to summarize: those related to the buyer size treatment (see Hypotheses 1 – 4) and those related to the market institution treatment (see Hypotheses 5 – 8). Broadly, the results show that a) large buyers exert more market power than small buyers and thereby capture a greater proportion of the gains from trade, and b) better quality and more cost-efficient abatement projects are traded in the $MANA$ market institution.

Credit prices were significantly higher in the small buyer markets, which indicates that large buyers were better able to push prices downward towards monopsony levels in a refutation of Hypothesis 1. Lower credit prices were instrumental in ensuring that large buyers got a significantly greater proportion of the gains from trade than the small buyers; a result that refutes Hypothesis 2.

Markets were parameterized such that profit differentials existed between monopsony and competitive equilibrium outcomes. This differential was larger in the large buyer markets. We may think of the profit differential as a "signal" to cooperate and generate the profitable monopsony outcome instead of compete and reach the competitive equilibrium outcome. Extending the analogy, the signal to cooperate is stronger in the large buyer market: the incentive to collaborate would be bigger for the large buyers given the greater profit differential. It is reasonable to assume that buyer awareness of the need to collaborate would be directly proportional to signal strength. And to assume that buyers who want to collaborate are likelier to do so, thereby increasing their profits. Patterns in credit price and capture of the gains from trade support this

chain of reasoning. It seems that large buyers did indeed cooperation tacitly, pushing prices down and generating higher profits than the smaller buyers. We infer that signal strength significantly affected the degree of cooperation between buyers.

Evidence on the aggregate quantity of credits traded and the number of successful trades indicates that trade quantity was closer to competitive equilibrium levels in large buyer markets than in small buyer markets, which supports the primary claim of Hypothesis 3. Hence prices were relatively closer to monopsony levels, but traded quantities were closer to competitive levels in large buyer markets. These two factors combined to cause the higher earnings of the large buyers. The low degree of quantity distortion also implied higher market trading efficiency in the large buyer market. Indeed the results did indicate significantly greater efficiency in the large buyer treatment, supporting Hypothesis 4.

The experimental results support Hypothesis 5, which states that abatement quality will be higher in the $MANA$ market than in the TR market. First, the POA metric of the market's physical efficiency, which measures abatement as a fraction of optimal abatement, shows the above statement to be true. Second, we find that mean abatement quality, as measured by $\bar{q}^M$, is significantly higher in the $MANA$ market. In the TR market only mean abatement is important, which as anticipated by Hypothesis 7 implies that mean abatement will be higher in the TR market. The results validate Hypothesis 7. The results also validate Hypothesis 6, which states that total abatement variability is lower in the $MANA$ market. The evidence suggests that the mean and variance of abatement is lower in the $MANA$ market. The difference between mean abatement in the TR market and mean abatement in the $MANA$ market is less than

the difference in the variance of the two markets. Since abatement quality is measured as $E-V$ these two facts together imply that abatement quality on the whole is higher in the $MANA$ market. The final important result on abatement quality is that abatements are more cost effectively obtained in the $MANA$ market, which validates Hypothesis 8. Since the *raison d'etre* of emissions trading markets is to create quality abatements cost effectively, the experimental results strongly support the $MANA$ market at the expense of the TR market.

Finally we summarize the results on the evolution of market power and efficiency within periods. Trading efficiency increases within periods and decreases across periods. The Price Index of Monopsony Trading Effectiveness on the other hand decreases within periods and increases across periods, which implies that time trends in buyers' market power parallel those in trading efficiency. Buyers' market power increases within periods and decreases across periods.

In conjunction, these results imply that as sellers reduce buyers' market power and prices move towards the competitive equilibrium level, market efficiency declines. This seemingly contrary outcome – where price movement towards competitive levels is accompanied by falling market efficiency – is explained by the process through which sellers combat buyers' market power. The trading pattern is for sellers in earlier periods to make deals even when their earning were very low. The basic imperative was to trade irrespective of earnings. In a low risk trading strategy sellers kept offer prices close to cost, thereby maximizing the probability of trade, but for low expected earnings. In later periods though, some sellers hold out until buyers raise their bids in a riskier trading strategy. These sellers keep their offer prices significantly above cost, thereby lowering

their probability of trade, but increasing earnings for all if they do manage to trade at a higher market price. If buyers raise bid prices to match the higher offer, market prices increase correspondingly, and raise profits for all sellers. Market price is determined by the marginal trade. A high market price requires a single seller with a high offer price and a single buyer to match with that offer. Note that as long as one seller offers high the others have the opportunity to free-ride by offering close to cost. The probability of the profitable trade is the same for the free-riders as for the high offerer. And, by bidding low the free-riders have a positive probability of earning some positive amount if the profitable trade does not occur. The high offerer, in contrast, earns nothing if her offer is not matched. The evidence shows that in later periods some sellers behave altruistically by taking on all the risk of not trading, in the hope of increasing profits for all sellers. We infer the existence of this process of holding out in later periods from the record of successful trades n. Holding the round constant, the number of successful trades is lower in the later periods, even though the subjects are more experienced. Analysis of offer patterns confirms that the drop in trade quantity is because some sellers keep offer prices high and remain unmatched.

Appendix A

A.1 Restricted Profit Functions

The following discussion draws heavily on results in Graff Zivin and Small (2003). Assume that point source j has output $x_j \geq 0$, which is accompanied by emissions $e_j \geq 0$. The output is produced by using the input vector $\boldsymbol{z}_j \in \mathbb{R}^n_+$. The production technology is described by $x_j = f(\boldsymbol{z}_j; e_j),\ \partial x_j/\partial z_j > 0 \forall\ z_j \in \boldsymbol{z}_j$. Output is increasing in each input and the the production function is assumed continuous and twice differentiable. Furthermore, production requires at least some amount of emission, $f(\boldsymbol{z}_j, 0) = 0$. Now, for any mix of inputs $\boldsymbol{z}_j$ let x_j increase until a certain level, after which diminishing returns set in. Hence, for any input mix, there is a certain maximum output. This maximum output is accompanied by the emission level $e^*_j(\boldsymbol{z}_j) = \arg\max_e f(\boldsymbol{z}_j, e_j)$. Thus for any given $\boldsymbol{z}_j$, $\partial x_j/\partial e_j > 0$ when $0 < e_j < e^*_j(\boldsymbol{z}_j)$ and $\partial x_j/\partial e_j = 0$ when $e_j = e^*_j(\boldsymbol{z}_j)$.

Let output price be r and let the input price vector be $\boldsymbol{w} \in \mathbb{R}^n_+$. Point source j's profit function is $\xi_j(\boldsymbol{z}_j, e_j) = rf(\boldsymbol{z}_j, e_j) - \boldsymbol{w}\boldsymbol{z}_j$. Now we know that j's emissions are constrained such that $e \leq e^*_j(\boldsymbol{z}_j))$. Given this information, the constrained profit maximization problem yields the following first order conditions: $\partial\xi_j/\partial z_j = r\partial x_j/\partial z_j - w_j = 0\ \forall\ z_j \in \boldsymbol{z}_j$. These first order conditions can be solved to yield an optimal $\boldsymbol{z}_j = \boldsymbol{z}^*_j(e_j)$. This solution to the constrained maximization problem is used to derive an expression for the restricted profit function by substituting the solution back into the

profit function ξ_j,

$$\pi_j(e_j) = \xi_j(\boldsymbol{z}_j^*(e_j), e_j) = rf(\boldsymbol{z}_j^*(e_j), e_j) - \boldsymbol{w}\boldsymbol{z}_j^*(e_j) \tag{A.1}$$

The restricted profit function for nonpoint sources is derived analogously through the substitution of (μ_i, σ_i^2) for e_j as arguments.

A.2 Profit Maximizing Outcome for Point Sources in MANA Market

The Lagrangean is of the form:

$$\begin{aligned} p\left[a_j - \sum_k \gamma_{kj} a_k\right] - [\pi_i^0 - \pi_j(a_j)] - \sum_i \gamma_{ij} q_i(E_i(\mu_i), V_i(\sigma_i^2)) \\ +\lambda_j \left[\sum_k \gamma_{kj} a_k + \sum_i \gamma_{ij} E_i - \sqrt{\frac{\sum_i \gamma_{ij}^2 V_i}{\beta}}\right] \end{aligned} \tag{A.2}$$

The point source j's control variables are $\{\gamma_{kj}, \gamma_{ij}\}_{i,k}$ and a_j. The Kuhn-Tucker conditions are

$$(\lambda_j - p)a_k = 0 \quad \forall\, k \tag{A.3}$$

$$p + \frac{\partial \pi_j}{\partial a_j} + (\lambda_j - p)\gamma_{jj} = 0 \tag{A.4}$$

$$-q_i(E_i(\mu_i), V_i(\sigma_i^2)) + \lambda_j \left(E_i(\mu_i) - \frac{\gamma_{ij} V_i(\sigma_i^2)}{\sqrt{\beta \sum_i \gamma_{ij}^2 V_i(\sigma_i^2)}}\right) = 0 \quad \forall\, i \tag{A.5}$$

$$\sum_k \gamma_{kj} a_k + \sum_i \gamma_{ij} E_i - \sqrt{\frac{\sum_i \gamma_{ij}^2 V_i}{\beta}} \geq 0 \tag{A.6}$$

$$\lambda_j \left(\sum_k \gamma_{kj} a_k + \sum_i \gamma_{ij} E_i - \sqrt{\frac{\sum_i \gamma_{ij}^2 V_i}{\beta}}\right) = 0 \tag{A.7}$$

From (A.3) it follows that $\lambda_j = p$. Substituting this result into (A.4) and using the definition of a_j yields (2.7). Equation (2.8) is obtained by substituting $\lambda_j = p$ into (A.5). Equation (A.6) is j's allowance constraint and (A.7) is the complementary slackness condition.

Bibliography

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Kluwer Academic Publishers, Dordrecht.

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Monographs in Statistics and Applied Probability 101. Chapman & Hall / CRC.

Baranzini, A. and Ramirez, J. (2005). Paying for quietness: The impact of noise on geneva rents. *Urban Studies*, 42:633–646.

Beavis, B. and Walker, M. (1983). Achieving environmental standards with stochastic discharges. *Journal of Environmental Economics and Management*, 10:103–111.

Ben-David, S., Brookshire, D., Burness, S., McKee, M., and Schmidt, C. (1999). Heterogeneity, irreversible production choices, and efficiency in emission permit markets. *Journal of Environmental Economics and Management*, 38:176–194.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons.

Beron, K., Hanson, Y., Murdoch, J., and Thayer, M. (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*, chapter Hedonic Price Functions and Spatial Dependence: Implications for the Demnd for Urban Air Quality, pages 267–281. Advances in Spatial Science. Springer-Verlag, Berlin.

Bossaerts, P., Fine, L., and Ledyard, J. (2002). Inducing liquidity in thin financial markets through combined-value trading mechanisms. *European Economic Review*, 46:1671–1695.

Brasington, D. and Hite, D. (2005). Demand for environmental quality: A spatial hedonic analysis. *Regional Science and Urban Economics*, 35:57–82.

Breetz, H., Fisher-Vanden, K., Garzon, L., Jacobs, H., Kroetz, K., and Terry, R. (2004). Water quality trading and offset initiatives in the us: A comprehensive survey. Available online at http://www.dartmouth.edu/ %7Ekfv/waterqualitytradingdatabase.pdf.

Cabe, R. and Herriges, J. (1992). The regulation of non-point-source pollution under imperfect and asymmetric information. *Journal of Environmental Economics and Management*, 22(2):134–146.

Carlèn, B. (2003). Market power in international carbon emissions trading: A laboratory test. Cambridge, MA. MIT Joint Program on the Science and Policy of Global Change. Report no. 96.

Carlson, D. and Sholtz, A. (1994). Designing pollution market instruments: Cases of uncertainty. *Contemporary Economic Policy*, 12:114–125.

Carpentier, C., Bosch, D., and Batie, S. (1998). Using spatial information to reduce costs of controlling agricultural nonpoint source pollution. *Agricultural and Resource Economics Review*, 27(1):72–84.

Carriazo Osorio, F. (2007). *Measuring Urban Amenities and Disamenities: A Spatial Hedonic Analysis in Bogotá, Colombia*. PhD thesis, The Pennsylvania State University.

Cason, T., Elliott, S., and van Boening, M. (1999). *Research in Experimental Economics*, volume 7, chapter Speculation in Experimental Markets for Emissions Permits. JAI Press, Stanford, CA.

Cason, T. and Gangadharan, L. (2003). Transactions costs in tradable permit markets: An experimental study of pollution market designs. *Journal of Regulatory Economics*, 23(2):145–165.

Cason, T. and Gangadharan, L. (2005). A laboratory comparison of uniform and discriminative price auctions for reducing non-point source pollution. *Land Economics*, 81(1):51–70.

Cason, T. and Gangadharan, L. (2006). Emissions variability in tradable permit markets with imperfect enforcement and banking. *Journal of Economic Behavior and Organization*, 61:199–216.

Cason, T., Gangadharan, L., and Duke, C. (2003). Market power in tradable emission markets: A laboratory testbed for emission trading in port phillip bay, victoria. *Ecological Economics*, 46(3):469–491.

Cason, T. and Plott, C. (1996). Epa's new emissions trading mechanism: A laboratory evaluation. *Journal of Environmental Economics and Management*, 30:133–160.

Clapp, J., Kim, H.-J., and Gelfand, A. (2002). Predicting spatial patterns of house prices using lpr and bayesian smoothing. *Real Estate Economics*, 30:505–532.

Cochard, F., Willinger, M., and Xepapadeas, A. (2005). Efficiency of nonpoint source pollution instruments: An experimental study. *Environmental and Resource Economics*, 30:393–422.

Cohen, J. and Coughlin, C. (2007). Spatial hedonic models of airport noise, proximity and house prices. Working Paper 2006-026C, Federal Reserve Bank of St. Louis.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, 17:563–586.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, revised edition.

Crocker, T. (1966). *The Economics of Air Pollution*, chapter The Structuring of Atmospheric Pollution Control Systems. W.W. Norton & Co.

Cronshaw, M. and Brown-Kruse, J. (1996). Regulated firms in pollution permit markets with banking. *Journal of Regulatory Economics*, 9(2):179–189.

Dales, J. (1968). *Pollution, Property and Prices*. University of Toronto Press, Toronto, Canada.

Davies, J. and Mazurek, J. (1998). *Pollution Control in the United States: Evaluation the System*. RFF Press, illustrated edition.

DeCoursey, D. (1985). Mathematical models for nonpoint water pollution control. *Journal of Soil and Water Conservation*, 40:508–513.

Economides, N. and Schwartz, R. (1995). Electronic call market trading. *Journal of Portfolio Management*, 21:10–18.

Edwards, W., Lindman, H., and Savage, L. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70:193–242.

Feitelson, E., Hurd, R., and Mudge, R. (1996). The impact of airport noise on willingness to pay for residences. *Transportation Research D*, 1:1–14.

Finley, A., Banerjee, S., and Carlin, B. (2007). spbayes: An r package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software*, 19(4).

Fisher, R. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fisher, R. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Franciosi, R., Issac, R., Pingry, D., and Reynolds, S. (1992). Extension of research into marketable emissions permits. Technical report, Oak Ridge National Laboratory. Final Report.

Franciosi, R., Issac, R., Pingry, D., and Reynolds, S. (1993). An experimental investigation of the hahn-noll revenue neutral auction for emissions licences. *Journal of Environmental Economics and Management*, 24:1–24.

Freeman, A. (2003). *The Measurement of Environmental and Resource Values, Theory and Methods*. Resources for the Future, second edition.

Gelfand, A., Ecker, M., Knight, J., and Sirmans, C. (2004). The dynamics of location in house price. *Journal of Real Estate Finance and Economics*, 29:149–166.

Gelman, A., Roberts, G., and Gilks, W. (1996). *Bayesian Statistics 5*, chapter Efficient Metropolis Jumping Rules, pages 599–607. Oxford University Press, Oxford.

Geweke, J. (1992). *Bayesian Statistics 4*, chapter Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. Clarendon Press, Oxford, UK.

Ghosh, G. and Carriazo-Osorio, F. (2007). Bayesian and frequentist approaches to hedonic modeling in a geo-statistical framework. Portland, OR. Selected Paper presented at the American Agricultural Economics Association Annual Meetings.

Ghosh, G. S. and Shortle, J. S. (2008). Water quality trading market efficiency in a stochastic environment. Selected Paper presented at the AERE Sessions at the AAEA & ACCI Joint Annual Meetings, Orlando FL, July 27–29, 2008.

Godby, R., Mestelman, S., Muller, R., and Welland, J. (1997). Emission trading with shares and coupons when control over discharges is uncertain. *Journal of Environmental Economics and Management*, 32:359–381.

Gotway, C. and Cressie, N. (1993). Improved multivariate prediction under a general linear model. *Journal of Multivariate Analysis*, 45.

Graff Zivin, J. and Small, A. (2003). Risk sharing in coasean contracts. *Journal of Environmental Economics and Management*, 45:394–415.

Greene, W. (2003). *Econometric Analysis*. Pearson Education, 5th (low price) edition.

Griffith, D. (1988). *Advanced Spatial Statistics*, volume 12 of *Advanced Studies in Theoretical and Applied Econometrics*. Kluwer Academic Publishers, Boston.

Gujarati, D. (2003). *Basic Econometrics*. McGraw-Hill, 4th (intl.) edition.

Hahn, R. (1988). Promoting efficiency and equity through institutional design. *Policy Sciences*, 21:41–66.

Hahn, R. and Hester, G. (1989). Marketable permits: Lessons from theory and practice. *Ecology Law Quaterly*, 16:361–406.

Hahn, R. and Noll, R. (1982). *Reform of Environmental Regulation*, chapter Designing a Market for Tradable Emissions Permits, pages 119–146. Ballinger, Cambridge, MA.

Haining, R. (1978). A spatial model for high plains agriculture. *Annals of the Association of American Geographers*, 68:493–504.

Haining, R. (2003). *Sapatial Data Analysis: Theory and Practice*. Cambridge University Press, Cambridge, UK.

Harrington, W., Krupnick, A., and Peskin, H. (1985). Policies for nonpoint-source water pollution control. *Journal of Soil and Water Conservation*, 40:27–32.

Hawkins, D. and Cressie, N. (1984). Robust kriging — a proposal. *Journal of the International Association for Mathematical Geology*, 16:3–18.

Hays, W. and Winkler, R. (1971). *Statistics: Probability, Inference and Decision*. International Series in Decision Processes. Holt, Rinehart and Winston, Inc., New York.

Heidelberger, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31:1109–1144.

Hennessy, D. and Feng, H. (2008). When to penalize uncertain nonpoint emissions be penalized in a trading program? *American Journal of Agricultural Economics*, 90(1):249–255.

Horan, R. and Shortle, J. (2001). *Environmental Policies for Agricultural Pollution Control*, chapter Environmental Instruments for Agriculture. CAB International, Oxfordshire, U.K.

Horan, R. and Shortle, J. (2005). When two wrongs make a right: Second-best point-nonpoint trading ratios. *American Journal of Agricultural Economics*, 87:340–352.

Horan, R., Shortle, J., and Abler, D. (2002). Point-nonpoint nutrient trading in the susquehanna river basin. *Water Resources Research*, 38:8–1–8–13.

Horan, R. D. (2001). Differences in social and public risk perceptions and conflicting impacts on point / nonpoint trading ratios. *American Journal of Agricultural Economics*, 83:934–941.

Howson, C. and Urbach, P. (1991). Bayesian reasoning in science. *Nature*, 350:371–374.

Hung, M. and Shaw, D. (2005). A trading-ratio system for trading water pollution discharge permits. *Journal of Environmental Economics and Management*, 49:83–102.

Intriligator, M. (1971). *Mathematical Optimization and Economic Theory*. Prentice-Hall, Englewood Cliffs, NJ.

Isaac, R., Ramey, V., and Williams, A. (1984). The effects of market organization on conspiracies in restraint of trade. *Journal of Economic Behavior and Organization*, 5:191–222.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, Oxford, UK, 3rd edition.

Kampas, A. and White, B. (2003). Selecting permit allocation rules for agricultural pollution control: A bargaining solution. *Ecological Economics*, 47(2-3):135–147.

Kim, C., Phipps, T., and Anselin, L. (2003). Measuring the benefits of air quality improvements: A spatial hedonic approach. *Journal of Environmental Economics and Management*, 45:24–39.

Lake, I., Lovett, A., Bateman, I., and Day, B. (2000). Improving land compensation procedures via gis and hedonic pricing. *Environment and Planning C*, 18:681–696.

Ledyard, J. and Szakaly-Moore, K. (1994). Designing organizations for trading pollution rights. *Journal of Economic Behavior and Organization*, 25:167–196. revenue neutrality, market power.

Leggett, C. and Bockstael, N. (2000). Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management*, 39:121–144.

Lichtenberg, E. and Zilberman, D. (1988). Efficient regulation of environmental health risks. *The Quarterly Journal of Economics*, 103:167–178.

Malik, A., Letson, D., and Crutchfield, S. R. (1993). Point/nonpoint source trading of pollution abatement: Choosing the right trading ratio. *American Journal of Agricultural Economics*, 75:959–967.

Mardia, K. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 76:289–295.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory*. Oxford University Press, New York.

Matheron, G. (1962). Traite de geostatistique appliquee. *Memoires du Bureau de Recherches Geologiques et Minieres*, 1(24).

Montgomery, W. (1972). Markets in licences and efficient pollution control regimes. *Journal of Economic Theory*, 5:395–418.

Morgan, C. and Wolverton, A. (2005). Water quality trading in the united states. Working Paper #05-07, National Center for Environmental Economics.

Muller, R. and Mestelman, S. (1994). Emissions trading with shares and coupons. *Energy Journal*, 15:185–211.

Muller, R. and Mestelman, S. (1998). What have we learned from emissions trading experiments? *Managerial and Decision Economics*, 19(4/5):225–238.

Muller, R. A., Mestelman, S., Spraggon, J., and Godby, R. (2002). Can double auctions control monopoly and monopsony power in emissions trading markets? *Journal of Environmental Economics and Management*, 44:70–92.

National Research Council (2001). Assessing the TMDL approach to water quality management. Technical report, National Research Council, Washington, DC.

Negahban, B., Moss, C., Jones, J., Zhang, J., Boggess, W., and Campbell, K. (1994). Optimal field management for regional water quality planning. St. Joseph, MI. ASAE Paper No. 94-3553. Presented at Annual Winter Meeting of American Society of Agricultural Engineers.

Nelder, J. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7:308–313.

Nelson, A., Genereux, J., and Genereux, M. (1992). Price effects of landfills on house values. *Land Economics*, 68:359–365.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heterscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Neyman, J. and Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A*, 231:289–337.

PADEP (2006). Trading of nutrient and sediment reduction credits - policy and guidelines. Available online at http://www.depweb.state.pa.us/chesapeake/.

Palmquist, R., Roka, F., and Vukina, T. (1997). Hog operations, environmental effects and residential property values. *Land Economics*, 73:114–124.

Patterson, H. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika*, 58:545–554.

Plott, C. (1983). Externalities and corrective prices in experimental markets. *Economic Journal*, 93:106–127.

Poe, G., Schulze, W., Segerson, K., Suter, J., and Vossler, C. (2004). Exploring the performance of ambient-based policy instruments when nonpoint source polluters can cooperate. *American Journal of Agricultural Economics*, 86(5):1203–1210.

Poirier, D. (1988). Frequentist and subjectivist perspectives on the problems of model building in economics. *Journal of Economic Perspectives*, 2:121–144.

Poor, P., Boyle, K., Taylor, L., and Bouchard, R. (2001). Objective versus subjective measures of water clarity in hedonic property value models. *Land Economics*, 77:482–493.

Pope, J. (2005). Information, attention and revealed preference: Limits to the hedonic model. Job Market Paper, North Carolina State University.

Qiu, Z., Prato, T., and McCamley, F. (2001). Evaluating environmental risks using safety-first constraints. *American Journal of Agricultural Economics*, 83(2):402–413.

Rabotyagov, S., Feng, H., and Kling, C. (2006). Optimal design of permit markets with an *ex ante* pollution target. Working Paper 06-WP 430, Center for Agricultural and Rural Development, Iowa State University, Ames, IA.

Ready, R. and Abdalla, C. (2005). The amenity and disamenity impacts of agriculture: Estimates from a hedonic pricing model. *American Journal of Agricultural Economics*, 87:314–326.

Reckhow, K. (1994). Importance of scientific uncertainty in decision making. *Environmental Management*, 18:161–166.

Reckhow, K. (2003). On the need for uncertainty assessment for TMDL modeling and implementation. *Journal of Water Resources, Planning and Management*, 129:245–246.

Ribaudo, M., Heimlich, R., and Peters, M. (2005). Nitrogen sources and Gulf hypoxia: Potential for environmental credit trading. *Ecological Economics*, 52:159–168.

Ridker, R. and Henning, J. (1967). The determinants of residential property values with special reference to air pollution. *Review of Economics and Statistics*, 49:246–225.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *The Journal of Political Economy*, 82:34–55.

Schabenberger, O. and Gotway, C. (2005). *Statistical Methods for Spatial Data Analysis*. Texts in Statistical Science. Chapman and Hall / CRC, Boca Raton, FL.

Segerson, K. (1988). Uncertainty and incentives for nonpoint pollution control. *Journal of Environmental Economics and Management*, 15:87–98.

Shortle, J. (1987). Allocative implications of comparisons between the marginal costs of point and nonpoint source pollution abatement. *Northeastern Journal of Agricultural and Resource Economics*, 16:17–23.

Shortle, J. and Abler, D. (1994). *The Economics of Nonpoint Pollution Control: Theory and Issues*, chapter Incentives for Agricultural Nonpoint Pollution Control. Kluwer Academic Press, Dortrecht, The Netherlands.

Shortle, J. and Horan, R. (2001). The economics of nonpoint pollution control. *Journal of Economic Surveys*, 15:255–289.

Shortle, J. and Horan, R. (2008). The economics of water quality trading. *International Review of Environmental and Resource Economics*, 2(2):101–133.

Shortle, J. S. (1990). The allocative efficiency implication of water pollution abatement cost comparisons. *Water Resources Research*, 26:793–797.

Shortle, J. S. and Dunn, J. W. (1986). The relative efficiency of agricultural source water pollution control policies. *American Journal of Agricultural Economics*, 68:668–677.

Smith, V. and Deyak, T. (1975). Measuring the impact of air pollution on property values. *Journal of Regional Science*, 15:277–288.

Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72:923–955.

Spraggon, J. (2002). Exogenous targeting instruments as a solution to group moral hazards. *Journal of Public Economics*, 84:427–456.

Spraggon, J. (2004). Testing ambient pollution instruments with heterogeneous agents. *Journal of Environmental Economics and Management*, 48:837–856.

Stein, M. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Annals of Statistics*, 16:55–63.

Taylor, L. (2003). *A Primer on Nonmarket Valuation*, chapter The Hedonic Method, pages 331–393. The Economics of Non-Market Goods and Resources. Kluwer Academic Publishers.

Taylor, M., Sohngen, B., Randall, A., and Pushkarskaya, H. (2004). Group contracts for voluntary nonpoint source pollution reductions: Evidence from experimental auctions. *American Journal of Agricultural Economics*, 86(5):1196–1202.

Tiebout, C. (1956). A pure theory of local expenditure. *Journal of Political Economy*, 64:416–424.

Tietenberg, T. (1990). Economic instruments for environmental regulation. *Oxford Review of Economic Policy*, 6:17–33.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.

Upton, G. and Fingleton, B. (1985). *Spatial Data Analysis by Example: Point Pattern and Quantitative Data*, volume I of *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, Chichester, UK.

USEPA (2001). The national costs of the total maximum daily load program (draft report). Technical report, Office of Water, Washington DC, 841-D-01-003.

USEPA (2002). National water quality inventory, 2000 report. Technical report, Office of Water, Washington DC, EPA-841-R-02-001.

USEPA (2003). Water quality trading policy. Technical report, Office of Water, Washington DC. Available online at http://www.epa.gov/owow/watershed/trading/tradingpolicy.html.

Vossler, C., Poe, G., Schulze, W., and Segerson, K. (2006). Communication and incentive mechanisms based on group performance: An experimental study of nonpoint pollution control. *Economic Inquiry*, 44(4):599–613.

Weersink, A., Livernois, J., Shogren, J., and Shortle, J. (1998). Economic instruments and environmental policy in agriculture. *Canadian Public Policy*, 24:309–327.

Wets, R. (1983). *Mathematical Programming: The State of the Art*, chapter Stochastic Programming: Solution Techniques and Approximation Schemes, pages 566–603. Springer, Berlin.

Woodward, R. (2000). Discussant's comments on "market based solutions to environmental problems. Invited Paper Session at the Annual Meeting of the Southern Agricultural Economics Association.

Xepapadeas, A. (1999). *Handbook of Experimental Economics*, chapter Nonpoint Source Pollution. Edward Elgar, Cheltenham, UK.

Zabel, J. and Kiel, K. (2000). Estimating the demand for air quality in four u.s cities. *Land Economics*, 76:174–194.

Vita

Gaurav Somenath Ghosh

Education

The Pennsylvania State University University Park, Pennsylvania 2004–2009
Ph.D. in Agricultural, Environmental & Regional Economics, expected in August 2009
Areas of Specialization: Spatial Econometrics, Market Design, Experimental Economics

Jawaharlal Nehru University New Delhi, India 2002–2004
M.A. in Economics, graduated with a First Class in the top 10% of class

St. Xavier's College, Mumbai University Mumbai, India 1999–2002
B.A. in Economics & Statistics, graduated with a First Class in the top 10% of class

Research Experience

Doctoral Research The Pennsylvania State University 2004–2009
Dissertation Co-Advisors: Dr. James S. Shortle & Dr. Anthony M. Kwasnica

- Designed and tested pollution trading markets
- Designed incentive mechanisms for voluntary pollution management
- Designed and implemented surveys for valuation study
- Valued air pollution damages in Bogotá, Colombia
- Wrote Quarterly Reports for USDA-funded project

Summer Internship Economic Research Service, USDA 2005
Research Advisor: Jorge Fernandez-Cornejo

- Analyzed resource conserving technology adoption among US corn and soybean farms

Summer Internship The Energy and Resources Institute (TERI) 2003
Research Advisor: Divya Datt

- Reviewed literature on sustainable development indicators

Working Papers

Water Quality Trading when Pollution Discharges are Stochastic, with James Shortle

Bayesian and Classical Inference in the Context of Spatial Hedonic Modeling, with Fernando Carriazo

A Laboratory Experiment to Compare Two Market Institutions for Emissions Trading, with James Shortle & Anthony Kwasnica

Do Baseline Requirements hinder Trades in Water Quality Trading Programs? with Marc Ribaudo & James Shortle

Professional Affiliations

American Agricultural Economics Association (AAEA) 2007–Present
Association of Environmental and Resource Economists (AERE) 2008–Present
Gamma Sigma Delta, Honor Society of Agriculture 2007–Present