EVALUATING SAMPLING BIASES IN POLICY ANALYSIS OF ENVIRONMENTAL MARKETS

BY

HONGSHUANG LI

B.Econ., Renmin University of China, 2007

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Consumer Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Advisor:

Assistant Professor Nicholas Brozović

Acknowledgments

It is my pleasure to thank many people who made this thesis possible. I gratefully acknowledge my thesis advisor Professor Nicholas Brozović for his continuous support and constructive suggestions and comments. Professor Brozović offered me the opportunity to join his project and provided important direction and focus for this thesis.[1] I could not have completed this work without going to his class - Environmental Economics. My gratitude also goes to my thesis committee members Professor John Braden and Professor Alex Winter-Nelson. Without their help, I would not have had a chance to join this program. I also want to thank them for their valuable and insightful comments on my thesis.

I would like to thank my colleagues, particularly those in my office: Eeshani Kandpal and Yusuke Kuwayama gave me a hand with coding, Xiaolin Ren, Xiang Bi, Jebaraj Asirvatham, and Taro Mieno shared so many of their research experiences with me. Thanks to Amanda Palazzo's careful records of her work, I could easily continue my work on this project. I also want to thank Sean Wan and Xiaoli Liao for their encouragement and support, Dr. Hongyun Jin, who advised my undergraduate study, and all my friends at that time, without whom I would not have the idea to go to the US for graduate study. Finally, I thank my parents for all they are and all they have done for me.

[1] This work was supported in part by the National Science Foundation under award number EAR-0709735.

To my parents – for their unconditional love

Table of Contents

List of Tables

List of Figures

Introduction and Motivation

Due to increasing water demands and, in particular, environmental concerns, more restrictions on agricultural water use are being established or negotiated in many parts of the world. The implementation of new restrictions on agricultural water use affects both total and individual production, and the welfare costs of restrictions may be distributed unevenly among heterogeneous producers. When there are many affected parties, both Pigouvian taxes and bargaining solutions may be difficult to implement. As a result, cap-and-trade systems are often suggested as politically feasible solutions to environmental externalities. In a competitive permit market with zero transaction costs, tradable permits are the least cost way to reach any target amount of abatement (Hanley et al. 2002). The permit price equalizes marginal abatement costs (MAC) of all resource users. Since users can decide whether or not to participate the market, no users can be worse off from trading in the permit market.

Empirical studies from several active water markets in the United States, Australia, and South Africa generally support the theoretical result that water is reallocated from low-value uses to higher-value uses through trading (Chong and Sunding 2006). Gains from trading in water markets are driven by heterogeneity of traders. Thus, in order to estimate the potential cost savings from introducing a tradable permit system and associated permit price, it is necessary to have data on relevant characteristics of all potential market participants. However, research data on characteristics of agricultural water uses are often very limited, particularly in the case of groundwater users (groundwater is typically a private property right and its use is unreported).

It is generally impractical to carry out a census to collect information for every potential policy, so sampling methods are a common approach to obtain data for policy analysis. However, there are many issues with sample design. What kind of sampling strategies are most accurate or most effective for *ex ante* studies of environmental markets? At which level of disaggregation should data be collected? Which characteristics should the sample try to maximize the representativeness of? Few previous environmental economic studies consider these questions, and most existing studies are based on samples with sparse data coverage, and state general conclusions without validating the sampling strategies and sampling procedure they used. Based on the information in a sample, some researchers explore the potential economic impacts of alternative water management policies (Pujol et al. 2006, Schaible 1997). But a sample can reveal only part of the characteristics of a population. If the sample is not representative of the population, then welfare analysis based on this sample may also be biased. As the heterogeneity of underlying population increases, and particularly if a relatively small portion of outlying data points have large effects on the market structure or outcomes, then the small sample may lead to biases in welfare estimates and even larger biases when scaled back to population estimates. In this case, policy implications from samples may be implausible and cannot be generalized to other regions and situations.

This thesis attempts to address these questions by making use of a unique population dataset of irrigation wells from the Republican River Basin of Nebraska. These data allow me to use a Monte Carlo framework to evaluate the effectiveness of alternative sampling strategies for estimating the welfare impact of a cap-and-trade groundwater rights system. Simple random sampling, systematic sampling, and stratified sampling, all of which are used in environmental economic studies, are applied in the welfare analysis of basin-wide trading. Additionally, the original data of a well level are aggregated to a farm ownership level, and compared to the preceding sampling strategies to see if sampling units have a significant influence on the sampling outcomes.

I obtain several results. First, results suggest that sampling biases in the welfare analysis based on one draw can be very large. The estimates for either area irrigated by one well or the costs saved by trading permits have the potential to be biased upwards by more than 10 percent. Second, wells are better sampling units than farms. This result follows because information at a well level is more disaggregated than that at a farm level, and large farms tend to own higher proportion of large wells. Third, when data are strongly heterogeneous in the estimated values of abatement costs per acre, scaling sample estimates back to population estimates by acreage can lead to much larger biases than scaling by the number of wells or farms. The last result is that the biased estimates of permit price can lead to substantial changes in estimated market structure, and this results in the most significant bias reported. For example, the samples on average imply a market with half buyers and half sellers, while only 20 percent of the participants are actually estimated to be sellers in the population data. The results in this thesis can be generalized to other environmental markets involving choice of sampling methodology, and bring insights into both *ex ante* sampling design and *ex post* diagnostics for sampling results based on sparse data.

This thesis is laid out as follows. I start with a review of related literature in Section 2, and describe the institutional background of water management in the Republican River Basin in Nebraska in Section 3. Then I describe the unique population dataset being studied in Section 4. In Section 5, I explain sampling strategies and detailed sampling steps as well as scaling methods used. After discussing the sampling results in different sampling strategies in Section 6, this thesis ends with some conclusions and broader implications for sampling approaches in evaluating environmental markets in Section 7.

Background

There are three strands of literature related to the research on welfare analysis of environmental markets presented in this thesis. The first strand considers the importance of spatial targeting of environmental policies in heterogeneous watersheds. The second strand either analyzes or simulates water markets to explore potential reasons for their success or failure. The third strand focuses on the design of sampling strategies, usually through surveys, to analyze watershed-scale agricultural water use. I discuss each strand below.

Spatial Heterogeneity in Watershed Characteristics

There is a relatively large literature on how spatial heterogeneity in watershed physical characteristics, and particularly those associated with environmental externalities, influences policy outcomes. For example, Satti and Jacobs (2004) showed the importance of including soil heterogeneity to capture the water requirements of individual farms. Yang et al. (2003) suggested that conservation costs can be reduced if abatement standards are set for heterogeneous regions rather than uniformly assigned. Diao et al. (2005) analyzed a surface water irrigation system in Morocco to evaluate the potential welfare gain from a spatially heterogeneous water allocation in agriculture. Several studies have used simulated economic data together with geo-referenced physical data, often in a Geographic Information System (GIS), to capture the benefits of spatial targeting of policies in a watershed (Braden et al. 1989, Khanna et al. 2003, Yang et al. 2003, Ancev et al. 2006). To date, most studies that analyze heterogeneous watershed characteristics assume that the economic agents using watershed resources are homogeneous. In contrast, this study explicitly incorporates both heterogeneous physical and economic properties of agricultural water use.

Water Markets

Studies of water markets focus on existing or potential markets to understand what drives transactions, what are practical problems in current markets, and what issues may undermine cost-efficient market function in general. Trading of water is driven by heterogeneous marginal benefits of irrigation, or equivalently by heterogeneous marginal abatement costs for water use reductions. If all farms had the same value of the marginal product of water at the current allocation, there would be no reason for trade to occur. Markets move scarce water resources from less productive users to more productive users to produce higher total benefits. The difference in marginal benefits exists both between agricultural, industrial, and residential sectors and within irrigation water use among heterogeneous farms. For example, 20 percent of sales in California's 1991 Drought Water Bank and 26 percent of sales in the Colorado-Big Thompson project area traded within the agricultural sector (Chong and Sunding 2006). Pujol et al. (2006) examined the potential benefits of water trading between 60 farms from Spain and Italy, and confirmed that water markets with no transaction costs could improve the economic efficiency of irrigation water use in the studied area.

Tradable water permits are the least cost way to achieve a fixed amount of abatement, potentially replicating the social optimal allocations (Hanley et al. 2002). Suppose the regulator needs to reduce overall water use to E. In the absence of regulation, user i will use an amount e_i^0 of water. Water use reduction lowers profits, and for abatement α_i, the cost is $C(\alpha_i)$. Then the regulator's problem is to minimize the cost of attaining the aggregate water use E.

$$Min_{\alpha_i} \quad \sum_{i=1}^{N} C(\alpha_i) - \lambda \left[E - \sum_{i=1}^{N} (e_i^0 - \alpha_i) \right] \tag{1}$$

$$\frac{\partial L}{\partial \alpha_i} = \frac{\partial C}{\partial \alpha_i} - \lambda = 0 \tag{2}$$

If equation (2) has an interior solution and α_i^* is the socially optimal abatement for user i, the first order conditions are $\frac{\partial C}{\partial \alpha_i} = \frac{\partial C}{\partial \alpha_j} = \lambda$, for any i, j. Thus, equalizing marginal abatement costs across users is the least-cost way to achieve any water use reduction.

If trading of permits is allowed and frictionless, then the market will reach an equilibrium price. Each user adjusts their abatement to minimize their total costs, through both changes in production and permit trading. The market mechanism provides an incentive for all users to abate pollution at the same marginal cost, and therefore achieves social optimal allocation as in equation (1) and equation (2).

User i's problem is

$$Min_{\alpha_i} \quad C(\alpha_i) - p^T \alpha_i \tag{3}$$

$$\frac{\partial L}{\partial \alpha_i} = \frac{\partial C}{\partial \alpha_i} - p^T = 0 \tag{4}$$

The first order conditions for (4) are $\frac{\partial C}{\partial \alpha_i} = \frac{\partial C}{\partial \alpha_j} = p^T$, for any i, j.

Comparing equations (2) and (4), the only differences are the Lagrange multipliers λ and p^T. The former one is the shadow price of social pollution damage, i.e. the social marginal abatement cost (MAC), and the latter one is the equilibrium permit price. If $p^T = \lambda$, the social optimum can be replicated in a competitive permit market with zero transaction costs.

Currently, the transfer of water rights encompasses a range of options, including water-use options, water right priority exchanges, and water banks (Chong and Sunding 2006). The optimal sizing of water markets has received some attention in the literature. In an empirical study of water transactions in New Mexico, Colby et al. (1993) showed that market prices are highly correlated with the size of transaction and the geographic range within which trading is allowed. Jenkins et al. (2004) examined the water system in California, and found that regional and statewide water markets could significantly reduce water scarcity and improve the flexibility and economic performance of water allocation.

The market-clearing price provides the correct incentive for farms to make decisions about crop choices and technology adoption. Even if landowners with water rights do not have much incentive to invest in a new technology under a water regulation system, they may adopt the technology under a water market system if the added revenue from the new technology can cover their fixed costs (Boggess et al. 1993). Theoretically, water markets have the potential to improve social welfare when compared with standards or regulations, but there may be practical or political problems in implementation. Beside the mobile nature and difficulty in identifying a specific part of the water, concerns also arise about wealth redistribution, environmental externalities, effects on third-parties, transaction costs, and uncertainty (Brozović et al. 2002, Pujol et al. 2006). Thus, water markets are often limited in their geographical scope.

Sampling Methods in the Study of Water Markets

Analyzing the performance of a water market requires data about the available water amount, agricultural production, marginal abatement costs, and so forth. The coverage and reliability of these data are critical to the validity of research results. For example, consider a sample that is used to estimate the welfare impacts of a potential water market. This sample is first used in equations (1) and (2) to calculate λ, the permit price that yields the correct aggregate water use. Then the trading behavior of each user and their total costs are estimated based on the permit price. Because the price is endogenous to the particular distribution of marginal abatement costs in the sample, each sample from a population will produce different permit prices, market behavior, and cost estimates. Thus, if properties of the sample are very different to the population of potential traders then there may be two important biases. First, estimates of cost savings may be biased. Second, trading behavior (buying or selling) and quantities traded estimated from a sample may be different both on average and for individual traders than obtained using population data.

However, population data on agricultural water use are rarely seen in the literature, and most studies are based on small samples without proving or testing the validity or representativeness of those samples. In general, analyses depend either on project-funded surveys within small regions or on one of two U.S. nationwide surveys (Table 1).

One nationwide survey is the Agricultural Resource Management Survey (ARMS). The ARMS survey is conducted annually by the Economic Research Service (ERS) and the National Agricultural Statistics Service (NASS) of the USDA, and collects information about farm structure, farm sector finance, and land use from the 48 contiguous states. ARMS is a probability-weighted and stratified survey.[2] Between 8000 and 10,000 farms are selected each year from the existing NASS List Frame as well as the Area Frame, exclusive of the List Frame (Goodwin et al. 2003). The List Frame is stratified by commodity type and sales class, while the Area Frame is stratified by land use categories (Katchova and Miranda 2004). Target states and crops vary each year to satisfy the complex stratification laid out in the ARMS methodology. Every surveyed farm has a specific probability as one factor in the data, which reflects the number of farms with similar attributes in the entire population of U.S. farms represented by this farm .

[2]More details about ARMS can be found at http://www.ers.usda.gov/Briefing/ARMS (accessed July 2009).

The stratified nature of the ARMS data is utilized in a wide variety of research on issues such as agricultural production, agricultural finance, and technology adoption.[3] One limitation of ARMS data is that the target states and crops vary each year, so that panel data are not directly amenable to time series analysis (Morrison-Paul et al. 2004).

Another nationwide survey – the Farm and Ranch Irrigation Survey (FRIS) – provides detailed information of on-farm irrigation activities, by targeting farms and ranches in all 50 states of the U.S.[4] It has been conducted by NASS every five years since 1974, as a follow-on survey of the Census of Agriculture. FRIS targets farms reporting irrigated land in the preceding Census of Agriculture. A sample of 20,000 to 25,000 irrigators is selected and mailed a report form in each survey, to cover 7 percent of the reported irrigated acreage. The sampling frames are constructed at the state level, and then a stratified sample is selected independently from 50 state frames. The stratification varies among the states, according to the distribution of total irrigated acres in a specific state. FRIS is designed to sample heavily on larger farms. It has a certainty stratum of the major irrigators in each state, whose farms are selected with probability one. For example, the national sample size was 25,014 farms in the 2003 FRIS survey.[5] Out of the national sample, 1823 farms were assigned to the certainty strata, while the remaining 23,191 farms were systematically selected from the noncertainty strata. FRIS provides consistent data for government decision, policy and regulation analysis, as well as for economic research.

Studies that do not use ARMS or FRIS data highly depend on the accessibility and voluntary participation of farm owners (Table 1). For example, Pujol et al. (2006) used linear programming to simulate a potential water market between Spain and Italy. In their data, 60 farms in Spain were chosen by quota sampling. They first

[3]Bibliography of published journal articles applying ARMS data can be found at http://www.ers.usda.gov/Briefing /ARMS/morereadings.htm (accessed July 2009).

[4]FRIS data exclude some horticultural farms and institutional, research, and experimental farms. The farms in the excluded categories in FRIS (2003), for example, accounted for 11 percent of the total number of irrigators and 2 percent of the irrigated land reported in the 2002 Census.

[5]FRIS data can be found in USDA Census of Agriculture. For instance, the 2003 FRIS data can be found in the 2002 Census Publication. The online version is at http://www.agcensus.usda.gov/Publications/2002/FRIS/index.asp (accessed July 2009).

classified all existing farms by farm size, and then sampled each class proportionately. The data on the Italian side were secondary population data of all the farms in the studied area, with a total of 131 farms. Quota sampling in Spain combined with the population data in Italy are assumed (but not tested) to capture the major characteristics of the potential water market across the border. Thus, the plausibility of simulations and conclusions based on small samples is generally untested and potential policy biases have not been evaluated.

Instead of survey approaches, some studies compile data from various sources to increase sample coverage and reveal more information about the target population. To the best of my knowledge, only two studies have attempted to analyze the population of water users in a large watershed. Hendricks (2007) used a unique dataset including 5075 parcels from 25 counties in Western Kansas to estimate the response of irrigation demand to water price and energy prices. He excluded some parcels due to complications with water and soil data, and focused only on estimating the demand elasticity of irrigation water. Palazzo (2009) assembled and analyzed a population dataset of all the irrigation wells in the Nebraska portion of the Republican River Basin to evaluate the cost savings of groundwater trading under alternative schemes. This thesis builds on the same dataset used by Palazzo (2009), but extends it to include farm-level analysis in addition to well-level analysis. Moreover, Palazzo (2009) did not consider the potential of sampling frames to bias policy analysis, or the broader implications of such biases to water resource management. Such issues are the focus of this study.

Institutional Context

Because groundwater is generally considered to be private property, there exist litter well-level data on its use. Moreover, even in regions where groundwater use is metered and allocated, trading of groundwater rights is usually highly restricted. The Republican River Basin is shared by Nebraska, Colorado, and Kansas. In 1942, these three states agreed to the Republican River Compact, which determined how to share the Basin's water resources and gave specific water allocations for "beneficial consumptive use" to each state (Hinderlider et al. 1942). The introduction of center pivot and sprinkler systems in the 1950s stimulated a sharp increase in well drilling in the Republican River Basin. Farmers extended irrigation by groundwater to fields not suited for furrow or ditch irrigation. As a result, water extraction for irrigation by all states increased, and in 1998 Kansas sued Nebraska and Colorado, claiming the upstream states were not leaving enough water instream to satisfy the Compact requirements. After litigation and a Supreme Court decision in 2002, the three states agreed on a final stipulation and determined reduced groundwater pumping allocations (McKusick 2002).

In all, there are currently around 11,000 active irrigation wells spread across the Nebraska portion of Republican River Basin, under the jurisdictions of the Nebraska Legislature, the Nebraska Department of Natural Resources (DNR), and local Natural Resource Districts (NRDs). The four NRDs in the Nebraska portion of the watershed are the Upper Republican, Lower Republican, Middle Republican and Tri-Basin NRDs, defined according to the characteristics of natural resource issues in each area. In order to preserve instream flows, the Nebraska legislature passed L.B.

962 in 2004, which required the NRDs to declare well drilling moratoria, meter wells used for irrigation, and set groundwater pumping limitations as well as to certify groundwater-irrigated acreage. The four NRDs accomplished well metering and the certification of groundwater-irrigated acreage by 2004. The certified acreage was issued based on the history of groundwater-irrigated production, and in some cases required tax assessments as proof of historical use. The Upper Republican NRD, located westernmost of the four NRDs, has the least rainfall, so it also has the highest current water allocation, namely 13 inches per acre per year (Nebraska Department of Natural Resources and Upper Republican Natural Resource District 2008). Based on precipitation, the allocation in the Middle Republican NRD is 12 inches per acre per year, and 9 inches per acre per year in the easternmost Lower Republican NRD, which has the highest rainfall amount (Nebraska Department of Natural Resources and Middle Republican Natural Resource District 2008, Nebraska Department of Natural Resources and Lower Republican Natural Resource District 2008). Irrigation allocations in the Tri-Basin NRD vary by county from 9 to 11 inches per acre per year (Nebraska Department of Natural Resources and Tri-Basin Natural Resource District 2007). Details of well location and allocations in each NRD can be found in Figure 1 and Table A-1 in the Appendix.

Groundwater trading is only a transfer of pumping rights, not involving actual conveyance of groundwater. However, due to legal issues, trade is highly regulated and restricted. Currently, the Upper and Middle Republican and Tri-Basin NRDs allow for trading of groundwater rights, approved by the relevant NRD Board of Directors, within townships or distance zones. The Lower Republican NRD does not currently allow any water trading.[6]

[6]Table A-1 also shows the current regulatory framework for groundwater trading in the four NRDs.

Data Description

This thesis is based on the population dataset of irrigation wells in the Republican River Basin (called "RRB data" in the following) assembled by Palazzo (2009). It combines the Republican River Compact Administration (RRCA) dataset of certified acreage, the Nebraska DNR well information database, the State Soils Geographic database (STATSGO),[7] Nebraska Cooperative Extension evapotranspiration (ET) data as well as other agronomic and economic parameters from *WaterOptimizer*. *WaterOptimizer* is a Microsoft Excel decision support tool developed by University of Nebraska-Lincoln Extension (Martin et al. 2005). It can simultaneously choose acreage, cropping patterns, and required water for a single field using a nonlinear optimization algorithm (Palazzo 2009).

The RRB data contain the certified irrigation acreage, pumping information, and geographic characteristics for all irrigation wells, and yields and prices for eight alternative crops.[8] Reductions from unconstrained water use to the current allocations set by regulations result in losses in profits, i.e. abatement costs, for water users. If using a market-based solution to achieve the same total reduction in water use, the abatement costs should be no larger than those from the regulation-based solution. The cost savings of interest in this thesis are the differences in abatement costs of switching from the current regulation to a cap-and-trade system with the same total water use. As described in Palazzo (2009), the single field nonlinear optimization solved by *WaterOptimizer* has been rewritten in Matlab and can be used to generate marginal abatement cost curves for water use reductions for every

[7] Soil type from the STATSGO dataset for Nebraska is categorized into coarse, medium, and fine types for this study.
[8] Alfalfa, corn (dryland and irrigated land), beans, sorghum, soybeans, beets, wheat.

well. Using these curves, the cost-effective reallocation problem shown in equation (1) and equation (2) can be solved. Similarly, equilibrium permit prices, trading activity, and total abatement costs for each well are obtained, allowing the water market structure and function to be analyzed.

I also aggregate all the fields irrigated by wells into farms based on the unique owner identification numbers in the Nebraska DNR well database. Distributions of estimated cost savings obtained from the population data analysis before and after aggregation are compared in the Appendix (Figure A-1 and Figure A-2[9]). The certified acreage of a single irrigation well varies from 1 acre to 557.4 acres, but reported acreages for owners with unique ID numbers ("farms") vary from 1 acre up to 14.7 thousand acres. The four largest farms account for 6 percent of the estimated population cost savings from basin-wide trading, while the thirteen largest farms account for almost 10 percent of the total cost savings. It indicates that the irrigation area of a farm (called "farm size" in the following) is more heterogeneous than the irrigated area by a well (called "well size" in the following).

There are four NRDs in the Republican River Basin (Figure 1); the NRDs are responsible for implementation of state groundwater management policies. Summary statistics for each NRD can be found in Table 2 and Table 3. Since one farm may have fields and irrigation wells in more than one NRD, in the farm-based sampling, I count the first NRD in which a farmer registered for certified acreage as the NRD he or she belongs to. Hence, the total acreage of each NRD in farm-based sampling is slightly different from that on well-based, but the differences are less than 1 percent. The average well sizes decrease from the western most Upper Republican NRD to the easternmost Lower Republican NRD. The Tri-basin NRD, located north of the Lower

[9]Figure A-1 shows the distributions of cost savings on well-based in the upper graph, and on farm-based in the lower graph. Name the number of wells (farms) in population as $N_p^W (N_p^F)$. In order to make the N_p^W wells and N_p^F farms comparable in terms of frequency, I lengthened the vertical axis in the lower graph by N_p^W/N_p^F. Horizontally, the farm-based cost savings is divided by N_p^W/N_p^F, since one farm owns N_p^W/N_p^F wells on average. Then I took logarithm of both horizontal axes to avoid squeezing most data to the left corner by several extremely large values. After these adjustments, both distributions appear similar shapes on the parts left to the population mean of cost savings. However, the right parts of well-based distribution shows a sharply downhill and a short tail, but the farm-based distribution has more frequency cumulates around 9 as well as a long and thin tail. It is possible that one farm owns a large number of small wells and reallocates a chunk from some left bins in the upper graph to an isolated dot at the right tail in the lower graph. But the aggregation is not reversible, i.e. it is impossible for some frequency in the upper graph ends up in a bin more left in the lower graph.

Republican NRD, has the smallest number of wells, but relatively large average well sizes. Note that the average acreage per farm in the RRB data are only about half of that in the FRIS survey (Table 3). There are several possible reasons for this discrepancy. First as shown by the total acreages, recent regulators have reduced the irrigated acreage from what was surveyed in the latest FRIS (2003). Second, the borders of NRDs are not consistent with the borders of counties. Except for the Upper Republican NRD, the other three NRDs contain several partial counties, but FRIS only provides county level data which I used to calculate the mean acreage in FRIS. Finally, another possible problem is that aggregation from wells to farms in the RRB data depend on unique owner IDs. Thus if members of the same family register as separate wells for legal or estate purposes, they will be recorded as separate farms.

Sampling Strategies

In any study employing sampling, a valid conclusion requires the sample to mirror the key characteristics of the population. However, no single sampling strategy is appropriate for all situations and the sampling design is also influenced by the budget and time constraints of the project.

The reasons leading to unrepresentativeness of a sample can be classified as either sampling error or non-sampling error. The former error is a fluctuation of sample estimates among different samples targeting the same population. The latter one results solely from the way an observation is made. The literature employs different definitions for both groups of errors. In some cases, “sampling bias” is used for problems resulting from poor sampling design, distinguished from sampling error which is caused merely by randomness. Moreover, nonresponse to a survey is sometimes considered to be sampling bias since researchers assume it to be a failure in sampling design. However, nonresponse is sometimes treated as a typical non-sampling error by other researchers because it happens during the process of observation (A comparison of sampling errors and non-sampling errors can be found in Assael and Keon (1982)). In this thesis, I use the Matlab software package to generate repeated random samplings and compute relevant results (the Monte Carlo method), so that observations are free of non-sampling errors. Furthermore, alternative sampling strategies are examined through the same sampling procedure. In the following discussion, I use “sampling bias” to refer to the difference between sample estimates and population values.

Sampling strategies fall into two major groups: probabilistic sampling and non-probabilistic sampling. Probabilistic sampling includes simple random sampling, systematic sampling, and stratified sampling. The basic idea for probabilistic sampling is the equal and independent chance for any element in the population to be selected. Non-probabilistic sampling, on the other hand, does not choose elements randomly. For example, the certainty stratum in FRIS has a probability of one to be included into the sample, and the quota sampling in Pujol et al. (2006) involves subjective judgment.

This thesis applies simple random sampling, systematic sampling, and stratified sampling on wells and farms basin-wide, and also applies simple random sampling to NRD-wide data.[10] The contribution of this thesis is not solely in testing the accuracy and efficiency of these sampling strategies, but also in providing some insights for welfare analysis involving sampling methods when market price determination and market structure depend on the sample. For example, consider the case where the government intends to sample 5 percent of farms in the Republican River Basin to measure the cost savings of allowing trading permits. Since different samples include different wells, each sample constructs a different market structure and consequently ends up at a different equilibrium permit price. The same farm benefiting from selling permits can gain thousands of dollars in a sample full of buyers, but very little in a sample with too many sellers relative to the population. The role of the same farm may change from buyer to seller, or vice versa, in various samples.

Another critical issue in sampling approach is the unit of analysis, which should be highly related to the area over which an individual decision-maker has control and information (Nelson 2002). In this thesis, both the acres irrigated by one well and the total irrigated area belonging to one farm owner can be counted as the basic sampling unit. The former is referred to as well-based sampling while the latter is referred to as farm-based sampling. Farm-based sampling is applied by FRIS and frequently used in agricultural economics. The comparison between the results from well-based sampling and farm-based sampling can help us to understand the effects of sampling

[10] As discussed previously, current regulations are implemented and data are collected at the NRD level.

units and the appropriateness of conventional farm-based sampling. In farm-based sampling on the RRB data, once a farm is included into a sample, all the irrigated acres owned by this farm are counted into this sample. Hence the number of wells in a sample is fixed in well-based sampling but fluctuates in farm-based sampling. For both well-based and farm-based samples, the production and abatement decision is made for every single well. Trade of permits is allowed both on-farm and off-farm.

From a statistical perspective, we should expect very similar outcomes from samplings on the same dataset with either sampling methodology. However, as shown in the next section, the differences are fairly large, since the aggregation from well-based to farm-based changes the distribution of cost savings.[11] The changes in the distribution of cost savings and the size of sampling units lead to different efficiency and accuracy in the estimates from three sampling strategies of interest: random, systematic, and stratified samplings. The population benchmarks for all of these samplings are \$16.34 million of estimated annual cost savings from frictionless basin-wide trading starting from current water allocations, associated with an optimal permit price of \$9.19 per inch per acre.

Simple Random Sampling

In simple random sampling, each well is assigned an equal probability of being sampled, regardless of subgroups or characteristics. For well-based sampling, the universal probability is $1/N_p^W$, where N_p^W is the number of wells in the population. The subscripts "p" and "s" in the following analysis mean population and sample statistics, separately, whereas the superscripts "W" and "F" represent well-based sampling and farm-based sampling, respectively.

Define N_s^W as the number of wells contained in an r percent sample, where N_s^W is the nearest integer to $N_p^W \times r\%$. Water permits can be traded between these N_s^W wells. Each well owner maximizes their total profit from both agricultural production

[11] Figure A-1 and Figure A-2 present the distributions of cost savings and cumulative density functions of acres for both sampling units.

and permit trading. In a competitive permit market, the market clearing price equalizes the marginal abatement costs (MACs) among all the wells and saves costs for each well (equation (2) and (4)). Sellers are compensated for increasing abatement by more than their cost of additional abatements. Buyers can reach abatement targets more cheaply by buying permits than by abating themselves. Since both sellers and buyers can choose whether or not to enter the permit market, cost savings are always non-negative. Then I apply a Monte Carlo approach by repeating this random procedure 1000 times to get robust sampling results. Similarly, I assign every farm with equal probability $1/N_p^F$ in the farm-based random sampling. To get an r percent sample, $N_s^F = N_p^F \times r\%$ farms are randomly selected. All the wells belonging to these N_s^F farms comprise the water permit market. However, the number of wells included in one sample can fluctuate in a wide range, depending on the specific farms included. At the optimal permit price, the water market clears and all N_s^F farms obtain non-negative cost savings.

As the sample size is raised from 1 percent to 5 percent, with 1 percent increments, the estimates of cost-savings from permit markets tend to be closer to the population benchmarks with lower standard deviations. The trend in estimates can be found in the Appendix (Figure A-3). Denote the irrigated acres as A_j^W for well j, and A_k^F for farm k. The associated cost saving is then W_j^W for well j and W_k^F for farm k. In order to scale the sample estimates of cost savings back to population estimates, I multiply the total cost savings of a sample by scaling ratios. Two scaling ratios are considered in this thesis: one is a simple numeric scaling ratio N_p^W/N_s^W, and the other is the conventional area scaling ratio $\sum_{j\in p} A_j^W / \sum_{j\in s} A_j^W$.[12] On farm-based sampling, the analogous ratios are N_p^F/N_s^F and $\sum_{j\in p} A_j^F / \sum_{j\in s} A_j^F$. For well-based sampling, the numeric scaling ratio is a fixed number, but the area scaling ratio varies depending on the specific wells included in the sample.

Define $\tilde{w}_i$ as the estimated total cost savings for the i^{th} draw after adjustment by

[12] Scaling the total cost savings in a sample by the area is equivalent to scaling the cost saving for each well or farm in this sample by area: $\sum_{j\in S}\{\frac{W_j^W}{A_j^W} \times \frac{A_j^W}{\sum_{j\in s} A_j^W}\} \times \sum_{j\in p} A_j^W = \{\sum_{j\in S} W_j^W\} \times \frac{\sum_{j\in p} A_j^W}{\sum_{j\in s} A_j^W}$.

either scaling ratio, and p_i as the associated permit price, where $i = 1, 2, \ldots, 1000$. The average estimate in 1000 draws is $\hat{W} = \sum_{i=1}^{1000} \tilde{w}_i/1000$ for cost savings, and $\hat{P} = \sum_{i=1}^{1000} \tilde{p}_i/1000$ for permit price. Recall that the population cost savings are \$16.34 million at the price \$9.19 per inch per acre, so the bias in 1000 draws is $(\hat{W} - \$16.34 \times 10^6)/\$16.34 \times 10^6$ in cost savings, and $(\hat{P} - \$9.19)/\$9.19$ in permit prices.

Simple random sampling is the most basic sampling strategy, but the randomness does not necessarily provide a representative sample for the population. If the target population is highly heterogeneous and the sampling results of interest rely on a few extreme observations, then simple random sampling may lead to large biases by missing or oversampling these extreme values. Consider an extreme case of a homogeneous dataset, which contains N_p^W identical wells. Any sampling approaches would produce the same sample estimate with no bias and zero standard deviation, whereas the estimating biases for the RRB data can be greater than 40 percent as shown in Section 5. Another concern is the difficulty to accomplish a really random sample. The chance of selection for each element can be influenced by many practical problems (Wockell and Asher 1994). For example, a complete list of all the elements in a large population can be hard to obtain. Even with a complete list to sample on, the sample designer may not be able to control the response rate.

Systematic Sampling

The significant heterogeneity in the RRB data implies possible efficiency gains in other sampling strategies over simple random sampling. If data are sorted in terms of characteristics of interest, systematic sampling may overcome the drawback of unrepresentativeness in simple random sampling. In systematic sampling, selections are evenly distributed along the ordered elements, and therefore avoid oversampling or undersampling certain types of elements. After sorting the population according to a certain criterion, a k percent systematic sampling selects every $(\frac{100}{k})^{th}$ element from a random starting point in the population. One advantage of systematic sampling is that it guarantees certain draws from both the low and high ends of a distribution proportionately. Departures between systematic sampling and simple random sampling tell us which type of elements are undersampled by random sampling. Systematic sampling is usually combined with cluster sampling or stratified sampling approaches in economic surveys, such as the sampling in noncertainty strata in FRIS.[13]

In this thesis, I sorted all wells in terms of their certified irrigated acreage, which when multiplied by the water allocation assigned by NRDs is the upper boundary of total available water for each well before trading. Because of the linkage between certified irrigated acreage and potential cost savings as shown in Figure 2, the application of a systematic sampling methodology should enhance sampling efficiency.

To get a 5 percent well-based sample, every 20^{th} element is recruited into the sample from a random start. For example, elements 22, 42, 62, $\ldots$, 10902 are a unique sample. Dividing N_p^W, i.e. 10,908, by 20 leaves a remainder of 8, so there are 28 possible unique systematic samples in total. I randomly started at an integer between 1 and 28, and then proceeded with every 20^{th} well. In each draw, the sampled wells could trade their permits at the equilibrium price.

The scaling of sample estimates is the same as that in simple random sampling. Total cost savings in the i^{th} draw, w_i, are multiplied by numeric ratio or area ratio to get $\tilde{w}_i$. The overall bias in cost savings is compared with $\$16.34 \times 10^6$, and the bias in permit prices is compared with \$9.19.

For the farm-based sampling, I sorted all the N_p^F farms in terms of their total certified acres, and then started randomly to select every 20^{th} farm to get a 5 percent sample. There are 25 possible unique samples out of 4525 farms. Additionally, I ordered farm owners by owner IDs, which were issued in order of the owner's first registration of an irrigation well.[14] If the group of early well owners has significantly

[13]FRIS has certainty strata, where each farm is assigned probability one to be selected into the samples, and noncertainty strata, where each farm is assigned an equal probability less than one.

[14]As shown in Figure A-4 and Figure A-5 in the Appendix, there exists an inverse U shape in the trend of marginal abatement costs along the registry order of wells.

different cost savings in trading permits, because of their demographics or geographic locations, this systematic sampling by owner ID should spread the sample equally among all well owners, and therefore reduce the biases from oversampling any subgroup of well owners. Then, the farm-based sample estimates are scaled as before and compared with the population benchmarks.

Stratified Sampling

Stratified sampling categorizes the population data into strata based on one or more criteria, and then samples each stratum independently. Because of the independence, different sampling strategies can be applied to each stratum. An efficient stratification requires that most variability lies between strata, minimizing the variability within one stratum. In ARMS, strata were decided based on multiple criteria including crop types, agricultural sales size, and land use categories. FRIS chose irrigated acreage as the criterion for stratification (FRIS (1988)).

In order to satisfy the variability requirement and also be comparable to the stratification in FRIS, I stratified the RRB data into three strata in terms of the total certified acreage at a farm level. The boundaries of these three strata are [1 acre, 160 acres], (160 acres, 320 acres] and (320 acres, 14700 acres]. This stratification is referred to as "ST1" in the following discussion. Because the basic unit of land division in Nebraska is a quarter section, or 160 acres, the cutting points are set at 160 acres and 160×2 acres.[15] Notice that in the preceding systematic sampling I ordered wells in well-based sampling by certified acreage. But in stratified sampling, I classify wells according to the farm they belong to in well-based sampling. Thus, if a farm has 160 certified acres or less, all the wells in this farm belong to the "small" stratum. In this way, well-based sampling and farm-based sampling have the same stratification and are comparable. I also stratified the data by acreage per farm into four strata by [minimum, 25th percentile], (25th percentile, 50th percentile], (50th

[15] Summary statistics for each stratum are presented in Table A-2 in the Appendix.

percentile, 75th percentile] and (75th percentile, maximum], referred to as "ST2". Finally, I applied a third stratification [min, 30th percentile], (30th percentile, 60th percentile], (60th percentile, 1000 acres][16], and (1000 acres, maximum], referred to as "ST3", to examine the sensitivity of the RRB data to different stratifications.

Using the first stratification, a 5 percent stratified well-based sample contains 5 percent of wells in each stratum, i.e. $N_p^{W_s} \times 5\%$ wells in the small stratum, $N_p^{W_m} \times 5\%$ wells in the medium stratum, and $N_p^{W_l} \times 5\%$ wells in the large stratum. The total number of wells in the sample is $N_s^W = N_p^{W_s} \times 5\% + N_p^{W_m} \times 5\% + N_p^{W_l} \times 5\%$. Then these N_s^W wells are used to construct a permit market, and consequently reach an equilibrium permit price. This sample is repeated 1000 times, and the total cost savings and the permit price in the i^{th} draw are denoted as w_i and p_i, respectively. Stratified farm-based sampling uses the same strata. Five percent of farms are selected from each stratum, and all the wells owned by these 5 percent of farms are included in the sample. The scaling of sample estimates and the calculation of biases is the same as for the preceding sampling strategies.

Random Sampling in NRDs

Basin-wide sampling may miss valuable information about subgroups. In multi-stage sampling, simple random samplings are usually applied in subgroups to provide the information or characteristics of specific subgroups of interest. Both ARMS and FRIS use states as subgroups in the first stage. Some other surveys further divide the first-level subgroup into smaller subgroups. In this thesis, I use the four NRDs (Upper Republican, Middle Republican, Lower Republican and Tri-basin) in the RRB data as subgroups, since currently regulations are implemented at the NRD level (see Table A-1 in the Appendix). Based on the variability of precipitation and current water allocations between NRDs, we expect different cost savings on average in each NRD. Each NRD is sampled independently and scaled back to the

[16] 1000 acres are an important stratum boundary in FRIS. In RRB data, the farm with 1000 acres stands for 83th percentile in the population ranked by farm size.

NRD level. For example, to take a 5 percent well-based sample from Upper NRD, I define w_i^u as the total cost savings of N_s^W wells sampled in the i^{th} draw. Notice that the population space is now the Upper Republican NRD instead of the entire Republican River Basin. Define $\tilde{w}_i^u$ as the total cost savings for the i^{th} draw after adjustment by either number scaling ratio or area scaling ratio, and p_i^u as the optimal permit price, where i=1, 2,..., 1000. The average estimates among 1000 draws are $\hat{W}$ and $\hat{P}$. Since the benchmarks for the Upper Republican NRD are \$5.88 million of cost savings at the permit price \$12.16 per inch per acre, the overall bias in the estimated cost savings is $(\hat{W} - \$5.88 \times 10^6)/\$5.88 \times 10^6$, and $(\hat{P} - \$12.16)/\$12.16$ is the bias in estimated permit price.

Similarly, w_i^m, w_i^l, and w_i^t represent the cost savings in the i^{th} draw from the Middle Republican, Lower Republican, and Tri-basin NRDs respectively. Their sample estimates are comparable to the cost savings of NRD-wide trading. Notice that in NRD-wide simple random sampling, the population space is reduced to less than one third of that basin-wide, so larger biases on average are predicted than in the preceding simple random sampling.

Results and Discussion

The Monte Carlo procedure described in the preceding section generates 1000 draws for three sampling strategies for the basin-wide data as well as for simple random sampling for the NRD-wide data. All these samplings are equal probability samplings. The system or stratification added to simple random sampling may improve the efficiency, but the biases cannot be totally removed for the 5 percent samples considered. Four major results are observed from the Monte Carlo analysis. First, biases in estimated cost savings through implementation of the water market can be fairly large in any single draw. Second, biases in farm-based sampling are much larger than those in well-based sampling. Thus, aggregation from wells to farms actually enlarges the sampling biases. Third, scaling from sample estimates to population estimates can also introduce large biases when the selection of scaling ratio does not take into account heterogeneity in the data. Last, but importantly, the largest biases are found in estimation of the equilibrium market price. In this section, I will discuss these results and the intuition underlying them in detail.

The means, medians, and standard deviations of biases in 5 percent samples are presented in Table 4 for well-based sampling and Table 5 for farm-based sampling. Table 6 and Table 7 are results for NRD-wide simple random sampling on wells and farms, respectively. The sampling statistics for 1 percent to 4 percent simple random sampling are presented in the Appendix.[17] The biases in the median of 1000 estimates follow the biases in the means in each table. Since using the median does not necessarily produce a lower bias, and the biases in median are close to those in

[17]Table A-3 and Table A-4 show the means, median and standard deviations of the biases in numeric scaled and area scaled cost savings, as well as permit prices. Figure A-3 presents trends in means, medians, 25th percentiles, and 75th percentiles.

means, I only analyze the latter bias in the following discussion.

The Biases in One Draw Can Be Large

The estimated cost savings in population trading are known for every single well or farm, so I first sampled on these fixed cost savings *ex post.* The average bias in 1000 draws is 0.08 percent for well-based sampling and -0.76 percent for farm-based sampling, both scaled by the number of wells. If scaled by area, the average bias is 0.1 percent for well-based sampling, and still -0.76 percent for farm-based sampling. However, the price in each draw is endogenous to the permit market built upon wells sampled in that draw. The estimation of permit price generates much larger biases. For example, simple random sampling results in -0.21 percent biases for well-based sampling as in Table 4, and 2.78 percent for farm-based sampling as in Table 5. Both exceed the biases in *ex post* sampling more than two times.

All the draws in each sampling are graphed in Figures 3 through 8. Each dot represents one draw of a 5 percent sample. The diamond is the average cost saving per acre at average well size or farm size. The four contours from inside to outside are the 20th, 40th, 60th, and 80th percentiles of probability density for the observations respectively, based on a two dimensional kernel density.[18] In Figures 3, 5, 6, and 8, the locally weighted scatterplot smoothing (Lowess)[19] indicates the trend among all the draws.

As can be seen in Figures 3 through 8, both positive and negative biases for estimated cost savings are found, and for any draw, the bias can be greater than 10 percent. Systematic sampling and stratified sampling can reduce the biases in well sizes or farm sizes, but neither sampling methodology has a good control on estimated cost savings, which are of particular interest in welfare analysis. Even if a sample exactly resembles the population in terms of area (well size or farm size), the

[18] Kernel density, or Parzen window, is used as a nonparametric way to estimate the probability density function of a random variable. Using a sample, the kernel density estimation can extrapolate the data to the entire population (Li and Racine 2007).

[19] All the Lowess smoothers shown in this thesis use a bandwidth of 0.9.

estimated cost savings can be biased up to 10 percent. A major reason for these biases is the endogenous permit price which determines the participation in the market and how much each market participant can save; this will be discussed again in more detail at the end of this section.

In practice, 1000 draws made by a Monte Carlo process may not be feasible. If there is only one chance for a survey to collect data, how much bias will exist in this sample? For example, the stratified well-based sampling provides the lowest biases among all sampling methodologies, but the estimated cost saving per acre varies from \$11 to \$15 (Figure 5). As the kernel density shows, a relatively large proportion of draws have cost saving estimates with biases of more than 10 percent compared to the population data.

Larger Biases after Aggregation

Although the same sampling approaches are applied to the same dataset, but with different sampling units, the biases are significantly enlarged in farm-based sampling. One percent of farms, namely 45 farms, produces a 32.56 percent bias on average in 1000 draws (Table A-4 in the Appendix). Even when sample size is doubled to include 90 farms – a larger sample size than several studies (3, 6, 7, and 8 in Table 1) – the average bias is still as high as 8.26 percent. Therefore more than 90 farms are needed to estimate the RRB data, if we want to control the bias to less than 8 percent on average.[20]

From the first section in Table 8 and Table 9, we know that in simple random, systematic, and stratified samplings on basin-wide data, the sampled area contained by 5 percent of farms on average is very close to the area contained by 5 percent of wells. So what leads to the overall larger biases in farm-based sampling? The average biases include both positive and negative values in well-based sampling (depending on sampling methodology), but all biases are positive in farm-based sampling (Table

[20]The sampling results for 1 percent to 5 percent farm-based samplings are in Table A-4 in Appendix, and the comparison with well-based sampling is in Table A-5

5). In addition, the mode in the distribution of the 1000 samples moves to higher cost saving per acre than the population average value (Figures 6 and 8). Recall that the bias in the *ex post* sampling of cost savings on farm-based generates 0.76 percent negative biases on average. The average estimates among 1000 draws should be robust enough to mitigate biases driven by extremely small or large farms, but what causes the positive biases in all the farm-based samplings?

After aggregation, the distribution of farm sizes is more heterogeneous than that of well sizes, but within farms, the well sizes are relatively homogeneous. So most wells contained in large farms are also large wells, while small farms generally contain only one or two small wells. The relationship between well sizes and farm sizes can be found in the Appendix (Figure A-6). If one large farm is included in a sample, this sample actually recruits a number of large wells. A hypothetical example serves to illustrate the implications of the relationship between farm size and well size. Consider a dataset that contains 100 small wells belonging to 100 small farms, 100 medium wells belonging to 50 medium farms (2 wells per farm), and 100 large wells belonging to a single large farm. So there are 300 wells owned by 151 farms in total. In a 5 percent well-based sampling, i.e. selecting 15 wells, it is impossible to include all of the 100 large wells, so the sample contains only 15 out of these 100 large wells at most. However, in farm-based sampling, the inclusion of all 100 large wells, i.e. the single large farm, is possible. If selecting a 40 percent sample to compare between two possible cases, 120 wells are chosen in a well-based sampling. The probability of obtaining all 100 large wells in one well-based sample is $\frac{\binom{200}{20}\binom{100}{100}}{\binom{300}{120}} = \frac{200!120!}{300!}\frac{1}{20!} < \frac{1}{20!}$. The numerator is the possible ways of choosing 20 wells out of the 200 small or medium wells and choosing all the other 100 large wells. The denominator is all the possible ways of choosing 120 wells out of all the 300 wells. But in farm-based sampling which selects 60 farms, the probability of sampling all 100 large wells is the probability of including the only large farm as well as 59 other farms into the sample, which is $\frac{\binom{150}{59}\binom{1}{1}}{\binom{151}{60}} = \frac{60}{151}$, much larger than $\frac{1}{20!}$. In this case, the numerator is the possible ways of choosing 59 farms out of the 150 small or medium farms and also choosing the only large farm. The denominator is all the possible ways of choosing 60 farms out of all the 151 farms. Therefore, if the large wells are mainly owned by large farms, a farm-based sampling has a higher probability of recruiting large wells. Recall that every 5 percent well-based sampling selects 545 wells out of 10,908 wells in the population. After aggregating these 10,908 wells into 4525 farms, a 5 percent farm-based sampling selects 226 farms in each draw. Although the number of farms is fixed at 226, the number of wells included can vary from 400 to 700 in different draws. So each farm has equal probability of entering the sample, but the sample actually contains more large wells due to the clustering of large wells in large farms.

Once it is clear that farm-based sampling includes more large wells than well-based sampling, Figure 2 can be used to explain why this can bias the estimated cost savings upwards. The upper panel in Figure 2 shows the MAC per acre for all the wells. The horizontal line is the market-clearing permit price in population trading, and the diamond shows the average MAC per acre at the average well size computed from population data. Hence, all the dots above the horizontal line represent wells estimated to purchase permits in the population trading analysis, whereas dots below this line are wells estimated to sell out permits. Buyers can reduce their MAC to the equilibrium permit price, and this reduction represents the cost savings. Sellers make a profit by selling permits at the equilibrium price, because they can abate water use relatively cheaply. There are some owners of small wells selling all their permits and switching to dryland farming, because irrigation produces almost no increased profit on their land (dots along the horizontal axis in Figure 2). Conversely, large wells are generally associated with higher MACs per acre because they gain relatively more than small wells from irrigation, due to their higher land quality and lower fixed costs per acre of pumping water. The curve through the data is a nonparametric Lowess smoother, which shows the average value of MAC per acre. Note that total cost savings after trading per acre are a function of the distance from the dots to the horizontal line representing the permit price, so the Lowess smoother tells us that the cost savings per acre generally decrease as well size increases and then increase slowly as the well size goes beyond 90 acres, which indicates that small wells and large wells gain the most on a per acre basis. Comparing the upper panel and lower panel of Figure 2, we can see that the heterogeneity in MACs is lower for very small and very large wells. Since the total cost savings of a well is given by cost saving per acre times area, the small wells still gain little in total, but the cost saving per acre in large farms is amplified by hundreds, or even thousands of times when total cost savings are calculated. Hence, oversampling of large wells will result in a positive bias on average in sample estimates for these data.

The lower graph in Figure 2 shows the average MAC weighted by area for each farm. The diamond is the average MAC per acre weighted by area at the average farm size in the population data. For instance, consider a farm with two wells: one of 20 acres with MAC c_1, and the other of 30 acres with MAC c_2. Then the average MAC weighted by area is $\frac{20}{50}c_1 + \frac{30}{50}c_2$. As before, the Lowess smoother shows that average MACs are increasing in farm size. In this case, this pattern is a reflection of the increased average profitability of large farms in irrigated agriculture relative to small farms. Several of the largest farms are left out of the lower graph to avoid compressing most observations close to the vertical axis. All these large farms are net buyers of water permits.

Biases from Scaling

Although we cannot reduce the bias to zero, if the bias is small enough, such as 2.78 percent in a 5 percent simple random farm-based sampling, the estimates should be acceptable for policy analysis or other economic research. However, biases larger than 20 percent appear in two groups of estimates: one is the cost saving estimates scaled by area in NRD-wide well-based sampling, and the other is the cost saving estimates in NRD-wide farm-based sampling, scaled by either ratio. What drives these large biases? I will discuss the first group of biases below and address the second group in the following subsection.

Recall the scaling ratios I used to adjust sample estimates back to population estimates. One is N_p^W/N_s^W, which infers population estimates by cost saving per well and the number of wells in the samples. The other is $\sum_{j\in p} A_j^W / \sum_{j\in s} A_j^W$, which infers population estimates by cost saving per acre and the sampled acreage. Unless cost saving per well or per acre is homogeneous, neither scaling ratio necessarily produces more accurate estimates.

Figure 9 (Figure 10) shows the cost saving per well (acre) in four NRDs. The horizontal lines are equilibrium permit prices in NRD-wide trading, and the diamonds are the average cost saving per well (acre) at the average well size. Of the wells, 94 percent irrigate between 1 acre and 200 acres. In Figure 9, the cost saving per well in this range is very close to the average, so the estimates adjusted by the number of wells in the first column of Table 6 have biases from -0.8 percent to 1.9 percent, much smaller than the biases (from -4 percent to 27 percent) in the area-adjusted estimates listed in the second column.

The statistics from the last four columns in Table 8 can provide us some insights about the causes of these large biases. Although I used the same series of random numbers for all sampling strategies, the simple random sampling happened to oversample small wells in the Upper Republican and Tri-basin NRDs, but oversample large wells in the Middle Republican and Lower Republican NRDs. For example, the irrigated acreage contained in the 5 percent samples from the Upper Republican NRD is on average 20.84 percent less than 5 percent of the total irrigated acreage in the Upper Republican NRD, which is 22620 acres. Then the question is why the area-adjusted estimates have 27.3 percent positive bias, when the sampled area is biased by -20.84 percent?

To understand this question, we need to refer to Figure 10, where the estimated cost saving per acre for each well is plotted against well size. The northwest panel, for the Upper Republican NRD, shows the source of the negative bias. The Lowess smoother shows that on average, cost savings per acre are highest for the smallest wells and decline steeply up to around 120 acres. Therefore oversampling small wells results in the large positive bias in area-adjusted estimates. A similar reasoning applies to the Tri-basin NRD. On the other hand, the samples for the Lower Republican NRD include too many large wells, which produces a 19.59 percent positive bias in the sampled acreage. As shown in the southwest panel of Figure 10, most large wells in the Lower Republican NRD irrigate between 100 acres and 200 acres, where the average cost saving per acre is smaller than the average (shown by a diamond), so oversampling on large wells leads to a -14.6 percent bias in area-adjusted estimates. In the Middle Republican NRD, the biases in both sampled area and area adjusted cost savings are modest.

In sum, the conventional scaling method based on area can lead to large biases when the target values per unit of area are strongly heterogeneous.

Biases through Endogenous Permit Price Determination

So far I have explained the reason for the first group of large biases. Does this reasoning also work on the second group? The area contained in farm-based samples for each NRD is biased from -3 percent to 3 percent as presented in Table 9, which means that the samples on average represent the population in terms of area. Thus, large biases should result from reasons other than scaling method. One hint we can get is the large bias (up to 50 percent) in estimated permit prices, which did not occur in any of the other 5 percent samplings. So what drives up estimated permit prices?

Tables 10 through 12 list the estimated percentage of buyers and sellers in the population, and in well-based samples and farm-based samples of NRD-wide markets. From Table 11, we know that 5 percent random well-based sampling provides a representative sample for the population data of each NRD. The percentage of buyers or sellers in the samples matches that in population on average. However, in farm-based sampling NRD-wide, there are important changes in these ratios.[21] The percentage of sellers is doubled in the Upper Republican, Middle Republican and

[21] The percentage of buyers or sellers represents "off-farm" trading, net of "on-farm" trading, where a farmer reallocates water between his/her parcels of land irrigated by different wells.

Tri-basin NRDs, and increased by 7 percent in the Lower Republican NRD, which implies the market structures in the samples are considerably different from the population market.

Figure 11 shows the average MACs weighted by area for farms less than 4000 acres plotted against farm size for four NRDs (the equivalent figure including all farms is Figure A-7 in the Appendix). The horizontal line in each graph is the permit price calculated from population NRD data, so farms lying above (below) these lines are net buyers (sellers) in the NRD-wide markets, and their cost savings are a function of the distance between their MACs and the permit price. In the northwest graph for the Upper Republican NRD, most farms larger than 500 acres are buyers, while the large sellers are sparse. In a 5 percent sample, only a few or even none of these large sellers are included in the sample to satisfy the demand from buyers. If a sample happens to miss all of these large sellers, those buyers lying relatively far away from the horizontal line will pull up the permit price and some previous buyers whose MACs are now lower than the new price will switch to sell their permits. This is why there are 41 percent sellers in farm-based samples trading at a price 20.6 percent higher, compared with 18 percent sellers in the population of the Upper Republican NRD. Similar changes also occur in the Middle Republican and Tri-basin NRDs. In the Lower Republican NRD, the asymmetry between buyers and seller is not as large as in the other three NRDs, but large buyers still outnumber large sellers. On average, 10 buyers in each sample switch to sellers due to a 12.2 percent increase in estimated permit price.

The average MAC of a farm implies the role of a farm in the market. Although permits are traded among wells, a farm with heterogeneous irrigated parcels trades within itself as well as trading with other farms. For instance, if the average MAC is lower than the permit price, the net effect is that a farm would sell its permits off-farm. Since many farms have average MACs close to the permit price, they are very sensitive to small changes in permit price. A slight bias in sampling can thus lead to some farms switching from being net buyers to net sellers (or vice versa) and further pushes up (pulls down) the price. The permit prices are endogenous to the market, through which they have a leveraging effect on sampling biases.

Sensitivity to Systems and Strata

Systematic sampling does not necessarily produce better estimates than simple random sampling. In Table 4, the average bias in systematic sampling is larger than that in simple random sampling and stratified sampling. But, in Table 5, the systematic sampling employs a conventional area scaling system, farm sizes, and reduces the bias to the lowest level of all three estimates. I also sorted farms by their owner ID numbers, i.e. the registry orders of their first wells. This sampling system generates 0.48 percent bias in average cost savings scaled by well number, 0.21 percent bias in cost savings scaled by area, and 5.52 percent bias in permit price. The magnitude of all of these biases is smaller than the magnitude of biases while ordering data by area. Therefore, farm vintage is also a factor to be considered in sampling strategies, although the conventional area scaling system can control the biases to some extent.

In farm-based stratified sampling, three stratifications are applied on the RRB data. The sampling outcomes in Table 5 employ "ST1":[1 acre, 160 acres], (160 acres, 320 acres], (320 acres, 14700 acres]. The average bias in cost savings is 7.13 percent if scaled by the number of wells, and 3.98 percent if scaled by area. In terms of magnitude, biases are less than -10.81 percent and -6.56 percent in the stratification into four quartiles ("ST2"), and -9.39 percent and -6.31 percent in "ST3": [1 acre, 30th percentile], (30th percentile, 60th percentile], (60th percentile, 1000 acres], and (1000 acres, 14700 acres]. The average bias in permit prices is -2.95 percent in "ST2" and -0.54 percent in "ST3", both of which are smaller in magnitude than that in "ST1". Thus, more strata do not necessarily imply better estimates. Although "ST1" has only three strata, it shows the lowest biases in estimated cost savings. Note that FRIS applies a stratification similar to "ST3" in this thesis, and scales estimates by area. Estimates using this methodology on average have 6.31 percent biases in the RRB data, which is equivalent to \$1.03 million in annual estimated cost savings. So, stratified sampling can potentially improve sampling accuracy and efficiency as it does in well-based sampling, but its performance also depends on the specific stratification as well as variation in the data.

Conclusion

Because of increasing conflicts over water use, agricultural water use is expected to see more restrictions in the near future. To understand the welfare impacts of alternative policies with sparse available data, sampling is commonly used to provide information for government decisions and economic studies. Therefore, it is critical to know whether the sample reproduces related characteristics of the population for welfare analysis. This thesis evaluates a potential environmental market using alternative sampling approaches to compare the effectiveness of each approach. Population data on 10,908 wells in the Republican River Basin of Nebraska (RRB) are used to estimate marginal abatement cost curves and hence how much money can be saved for each well through trading water permits instead of the current water allocation scheme. Simple random sampling, systematic sampling, and stratified sampling are examined using both well-level data and farm-level data. With Monte Carlo methods, I use 1000 draws of each sampling strategy for robust results.

First, sampling outcomes show that the biases in welfare analysis can be fairly large in a single draw. Second, aggregation from wells to farms increases the biases on average. Furthermore, scaling methods can significantly enlarge the sampling biases in the NRD-wide random sampling. Last, biases in estimated permit price can lead to large changes in estimated water market structures. This set of results provides general implications for the evaluation of environmental markets.

The target of any sampling strategy is to understand the population through only a part of the potentially available information. Thus if the results are biased, their application either locally or more generally may be invalid. Usually, the sample size does not exceed 5 percent in agricultural water use samples of the kind considered in this thesis. When researchers have a single 5 percent sample, rather than 1000 draws as in this thesis, how much bias can be expected *ex ante*? The answer depends on the sampling strategies, sample sizes, and scaling methods used. In this thesis, a single draw has the potential to contain large biases (up to 20 percent) in all estimates of interest. In practice, this implies a high probability for a survey to collect biased information and lead to incorrect analyses. Moreover, neither the conventional sampling unit (farms) or the conventional scaling method (scaling by acres irrigated by a well) applied on the population dataset in this thesis provides more accurate estimates, compared with alternatives. Instead, disaggregating the data from the farm level to the well level can improve the accuracy and the efficiency of estimates. Conventional scaling by area is only appropriate when the target values per acre are homogeneous. In the data considered in this thesis, the cost savings per acre change significantly at different well sizes or farm sizes. Scaling the sample estimates by the number of wells instead of the acres irrigated by a well can avoid biases of up to 27 percent in the NRD-wide sampling. The most important concern raised by this thesis is the extent to which a single sample can reproduce a market structure mirroring the market structure estimated for the population. Unfortunately, the bias in estimated permit price can result in substantial changes in estimated market structures.

In order to understand the data and correct biases in sampling, it is critical to figure out the underlying features that affect the estimated values to be analyzed. In the RRB data, the cost savings and permit prices of interest are a function of the area irrigated by a well or a farm, so the heterogeneity in cost savings per acre has a large impact on the selection of sampling strategies and scaling methods. If no information about the target basin is available, which means that there is no information on relevant heterogeneity of the underlying population, my research suggests that randomly sampling on wells (if possible) rather than farms will provide more accurate and more efficient estimates. This is important because several existing surveys of agricultural water use, such as the Farm and Ranch Irrigation

Survey, are based on a stratification at farm-level (for example, all papers except 3 and 4 in Table 1). If several sampling units are available, choosing the relatively homogeneous one would decrease the burden on sample size and *ex post* diagnostics.

When available data are restricted, *ex post* analysis may be used to diagnose bias to some extent. For instance, graphing estimated values along the most influential underlying characteristics in the data may highlight trends in the population data. Plotting estimated values and influential characteristics may show how sensitive the welfare estimates and market structures are to the sample chosen. For example, if most observations are clustered far away from the estimated permit price, the researcher does not need to worry about a switch in market behavior due to the bias in estimated permit price.

This thesis analyzes alternative sampling approaches for an irrigation water market. However, its consideration of the heterogeneity in data and leveraging effects of market prices on estimated market structure and welfare impacts can be further applied to other welfare analysis. I only focused on the amount of water use in irrigation, with one production technology in the same watershed, but producer heterogeneity was still found to be large. Other potential cap-and-trade markets, such as a carbon cap-and-trade market or a water quality market, are expected to show an even higher degree of heterogeneity in the underlying population, and thus even larger potential biases in welfare estimates from sampling. Moreover, an analysis of transaction costs may also further enlarge the biases in estimation. Another potential future approach is to apply Bayesian methods to offer detailed *ex post* diagnostic schemes for studies with small samples.

Tables and Figures

Table 1: Studies of Irrigation Water Use and Associated Survey Methods

	Author	Question	Data	Sampling	Method
1	Moore et al. (1994)	Demand response to water price	FRIS (1984) and FRIS (1988), 2442 farms in total	Survey	Econometrics
2	Schaible (1997)	Analyze water conservation policies	FRIS	Survey	Programming
3	Ise and Sunding (1998)	Who sells their water rights to the government	65 obs (30 sellers and 35 nonsellers) , 1/3 of the acreage in the Carson Division of the project.	Voluntary; from a telephone and mail survey	Econometrics
4	Gonzalez-Alvarez et al. (2006)	The response of irrigation water use to marginal costs	707 obs in 1999-2002, each obs is a field of one crop for one season. It accounts for 4% of total GA farmers.	Survey; survey design is based on irrigation source, crop type, technology and geographics	Econometrics
5	Koundouri et al. (2006)	Technology adoption	265 farms in Crete in 1995	Randomly selected; survey	Econometrics
6	Diwakara and Chandrakanth (2007)	A watershed development program	65 farms	Survey	Econometrics
7	Pujol et al. (2006)	How will water market improve the efficiency of water allocation	60 farms in Spain and 131 farms in Italy.	Quota sampling and accessibility; data from Spain is selected by quota sampling, based on their size; data from Italy is secondary data, 131 farms in a sub-area	Linear programming
8	Moore and Dinar (1995)	Are water and land quantity-rationed inputs?	79 farms in 1988	Accessibility; survey	Econometrics

Table 2: Summary Statistics of Certified Acreage in Each NRD, Well-based

	Number of Wells	Mean Acreage Per Well	Total Acreage	% of Total Acreage
Upper NRD	3183	142	452395	36.21%
Middle NRD	2876	107	307505	24.61%
Lower NRD	3320	94	313514	25.09%
Tri-basin NRD	1529	115	175907	14.08%
Total	10908	115	1249322	100.00%

Table 3: Summary Statistics of Certified Acreage in Each NRD, Farm-based

	Number of Farms	Mean Acreage Per Farm	Mean Acreage Per Farm in FRIS	Total Acreage	Total Acreage in FRIS	% of Total
Upper NRD	1194	385	775	459703	379505	36.80%
Middle NRD	1234	243	411	299551	396775	23.98%
Lower NRD	1358	226	392	306490	339952	24.53%
Tri-basin NRD	739	248	649	183579	549817	14.69%
Total	4525	276	525	1249322	1666049	100.00%

It is possible for one farm to own wells in more than one NRD. When aggregating wells into farms by owner IDs, farm owners are categorized by the first NRD in which they registered their wells. Therefore, the area in each NRD changes within 1 percent.
The mean acreage in FRIS is an area weighted mean.

Table 4: Five Percent Well-based Sampling on Basin-wide Trading

		Cost Saving (10^7) Scaled by the Number of Wells	Cost Saving (10^7) Scaled by Area	Permit Price
Population	Value	1.63	1.63	9.19
Random	Mean	-0.21%	-0.19%	0.03%
	Median	-0.07%	-0.03%	-0.06%
	std.	5.37%	5.06%	3.49%
Systematic	Mean	-0.36%	-0.24%	-0.41%
	Median	0.42%	0.58%	-0.06%
	std.	5.08%	5.09%	3.16%
Stratified	Mean	0.03%	0.03%	0.15%
	Median	0.19%	0.26%	0.02%
	std.	5.09%	5.04%	3.39%

The first row shows the population values in basin-wide trading. The following rows present the percentage of biases or standard deviations in 1000 draws.

Table 5: Five Percent Farm-based Sampling on Basin-wide Trading

		Cost Saving (10^7) Scaled by the Number of Wells	Cost Saving (10^7) Scaled by Area	Permit Price
Population	Value	1.63	1.63	9.19
Random	Mean	2.78%	3.00%	12.19%
	Median	1.88%	2.62%	12.61%
	std.	8.59%	7.41%	4.03%
Systematic	Mean	1.30%	0.66%	9.62%
	Median	0.29%	0.26%	9.74%
	std.	9.06%	7.03%	4.52%
Stratified	Mean	7.13%	3.98%	13.09%
	Median	3.66%	4.03%	13.12%
	std.	8.54%	7.53%	4.23%

Table 6: Five Percent Well-based Sampling on NRD-wide Trading

		Cost Saving (10^6) Scaled by the Number of Wells	Cost Saving (10^6) Scaled by Area	Permit Price
Upper NRD	Population	5.88	5.88	12.16
	Mean	0.74%	27.35%	-0.20%
	Median	0.32%	27.07%	-0.25%
	std.	9.60%	10.32%	3.88%
Middle NRD	Population	4.52	4.52	7.17
	Mean	0.80%	-4.20%	0.09%
	Median	0.29%	-4.60%	0.00%
	std.	11.26%	12.08%	9.54%
Lower NRD	Population	2.77	2.77	7.66
	Mean	1.91%	-14.63%	-1.01%
	Median	1.93%	-14.73%	-0.59%
	std.	102.16%	11.12%	7.53%
Tri-basin NRD	Population	1.00	1.00	9.18
	Mean	-0.84%	14.34%	-0.37%
	Median	-1.05%	14.43%	0.53%
	std.	19.38%	20.05%	8.24%

Population values are derived from NRD-wide trading. The three rows following each population value are percentage of biases or standard deviations in 1000 draws.

Table 7: Five Percent Farm-based Sampling on NRD-wide Trading

		Cost Saving (10^6) Scaled by the Number of Wells	Cost Saving (10^6) Scaled by Area	Permit Price
Upper NRD	Population	5.88	5.88	12.16
	Mean	43.56%	44.20%	20.63%
	Median	32.40%	32.49%	19.66%
	std.	29.91%	30.74%	7.96%
Middle NRD	Population	4.52	4.52	7.17
	Mean	34.08%	34.25%	49.50%
	Median	32.36%	30.10%	49.64%
	std.	21.01%	19.66%	11.37%
Lower NRD	Population	2.77	2.77	7.66
	Mean	6.69%	6.97%	12.15%
	Median	5.88%	6.48%	12.84%
	std.	13.13%	12.88%	8.02%
Tri-basin NRD	Population	1.00	1.00	9.18
	Mean	8.13%	8.33%	11.94%
	Median	6.13%	5.63%	14.15%
	std.	24.08%	24.62%	8.70%

Table 8: Bias in Sampled Area, Lift, and Yield for Well-based Sampling

		Random	Systematic	Stratified	Upper	Middle	Lower	Tri-basin
Area in sample	Population	62397	62397	62397	22620	15375	15676	8795
	Mean	-0.69%	-0.08%	0.38%	-20.84%	5.57%	19.59%	-13.51%
	Std.	10.99%	0.34%	1.47%	4.07%	4.18%	3.83%	6.23%
Lift	Population	147	147	147	135	162	131	180
	Mean	0.19%	-0.09%	0.08%	-0.08%	-0.07%	-0.14%	0.04%
	Std.	4.68%	1.49%	2.45%	4.09%	5.56%	4.57%	3.36%
Yield	Population	1045	1045	1045	1510	869	758	1028
	Mean	-0.12%	0.17%	0.06%	-0.06%	0.22%	-0.04%	-0.10%
	Std.	4.58%	1.75%	2.40%	3.66%	4.65%	3.67%	3.58%

In each section, the first row is the population value, the second row is the bias in the mean of estimates, and the third row is the standard deviations in the estimates.
Lift is well pumping water level (feet), and yield is well yield (gallons per minute).

Table 9: Bias in Sampled Area, Lift, and Yield for Farm-based Sampling

		Random	Systematic	Stratified	Upper	Middle	Lower	Tri-basin
Area in sample	Population	62397	62397	62397	22960	14961	15308	9169
	Mean	-0.51%	-1.48%	0.07%	-1.48%	3.07%	2.62%	-2.83%
	Std.	10.97%	11.22%	9.29%	22.94%	217.69%	12.01%	14.57%
Lift	Population	147	147	147	135	162	131	180
	Mean	0.19%	0.48%	0.31%	0.82%	0.13%	0.31%	0.02%
	Std.	4.68%	4.63%	4.59%	8.27%	9.75%	8.40%	5.32%
Yield	Population	1045	1045	1045	1510	869	758	1028
	Mean	-0.12%	0.10%	-0.19%	0.05%	0.17%	-0.11%	0.10%
	Std.	4.58%	5.75%	4.72%	5.67%	7.67%	6.12%	4.68%

Table 10: Buyers and Sellers in Population

	Sellers	%	Buyers	%	Outsiders	%	Sell All(%)	Permit Price	Total Trading Acre-inches
Basin	2452	23%	7337	67%	1119	10%	21%	9.19	2054166
Upper	571	18%	2555	80%	57	2%	17%	12.16	896824
Middle	643	22%	1742	61%	491	17%	19%	7.17	495033
Lower	705	21%	2070	62%	545	17%	20%	7.66	453315
Tri-basin	190	12%	1313	86%	26	2%	12%	9.19	173223

The first row shows the numbers and percentage of sellers, buyers, outsiders, and people selling out all permits, as well as permit prices and total traded amount in basin-wide trading. The following four rows shows analogous values in NRD-wide trading. "Outsiders" are nonparticipants to the permit market, since their marginal abatement cost is the same as permit price.

Table 11: Buyers and Sellers in Well-based Sampling

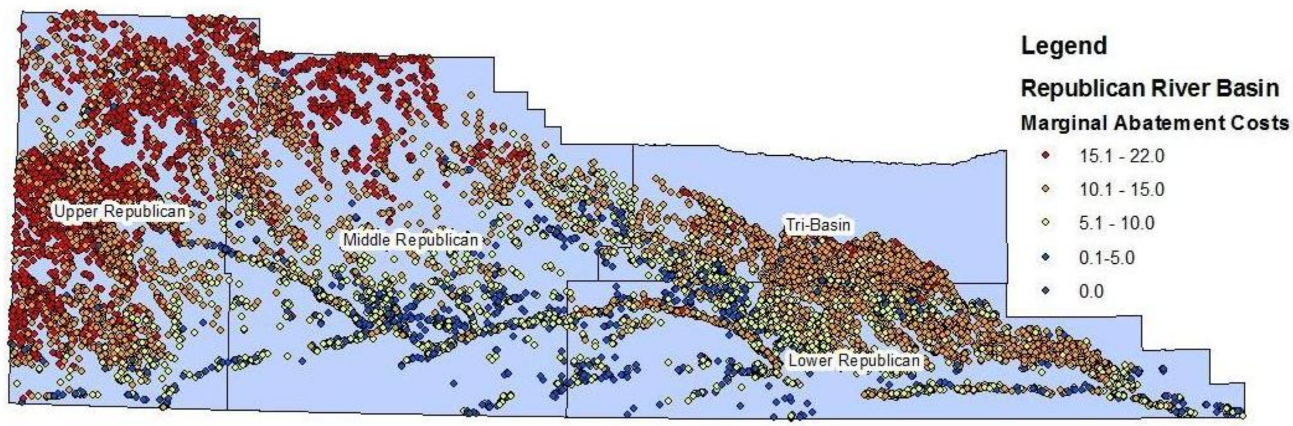
	Sellers	%	Buyers	%	Outsiders	%	Sell All(%)	Permit Price	Total Trading Acre-inches
Upper	29	18%	127	80%	3	2%	16%	12.14	45281
Middle	32	22%	87	60%	25	17%	19%	7.18	25166
Lower	35	21%	104	62%	27	16%	20%	7.59	23133
Tri-basin	10	13%	65	85%	1	2%	12%	9.15	8759

Table 12: Buyers and Sellers in Farm-based Sampling

	Sellers	%	Buyers	%	Outsiders	%	Sell All(%)	Permit Price	Total Acre-inches Trade
Upper	62	41%	94	57%	3	2%	40%	14.67	82198
Middle	60	43%	59	40%	24	17%	38%	10.73	38545
Lower	45	28%	94	56%	28	17%	26%	8.60	26000
Tri-basin	17	22%	59	76%	1	2%	20%	10.28	11661

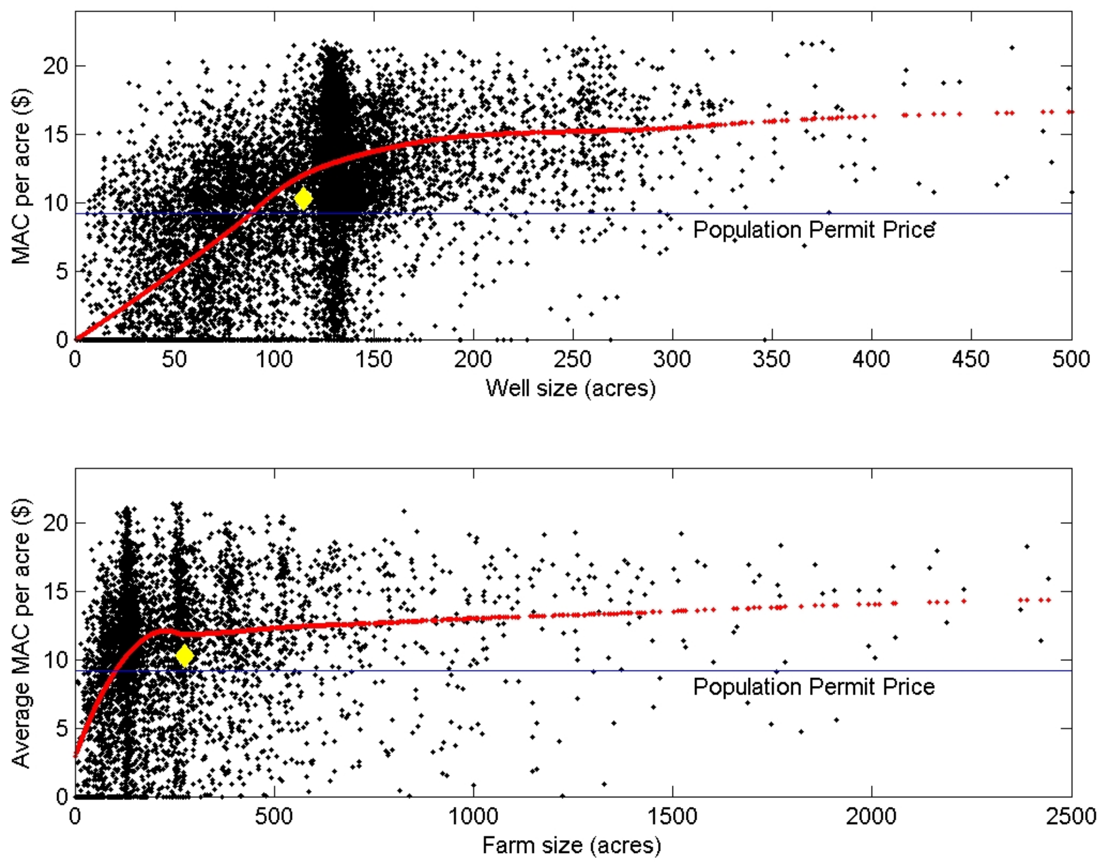
"Outsiders" are farms with no off-farm trading of water permits, since their average marginal abatement cost is the same as permit price. These outsiders may have on-farm trading, all of which are counted into the permit market in analysis.

Figure 1: Wells in the Republican River Basin (NE) with Certified Acreage



This figure is reproduced from Palazzo (2009). Each dot represents a well with certified irrigation acreage. It shows the marginal abatement costs in dollars per inch per acre of reducing water use at current water allocations, using *WaterOptimizer* (Martin et al. 2005). The current allocations are as follows: 13 inches per acre in Upper Republican NRD; 12 inches per acre in Middle Republican NRD; 9 inches per acre in Lower Republican NRD; 9-11 inches per acre in Tri-Basin NRD.

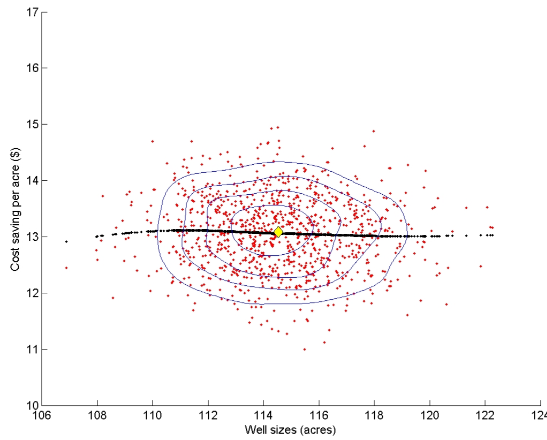
Figure 2: Marginal Abatement Costs as A Function of Well Size and Farm Size



In the upper panel, the marginal abatement cost per acre for each well are plotted against well sizes. The curve is Lowess smoother, while the horizontal line is the equilibrium permit price of population(basin-wide) trading. The diamond in the mean of marginal abatement costs at the average well size.

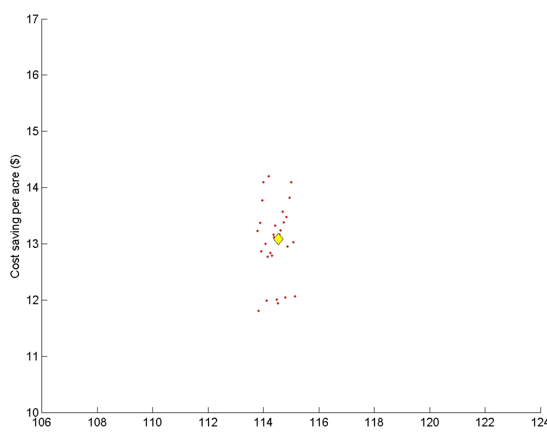
In the lower panel, the average marginal abatement cost per acre weighted by area for each farm are plotted against farm sizes. The Lowess smoother, horizontal line and the diamond are defined in the same way as in upper panel.

Figure 3: Random Samples for Well-based Sampling



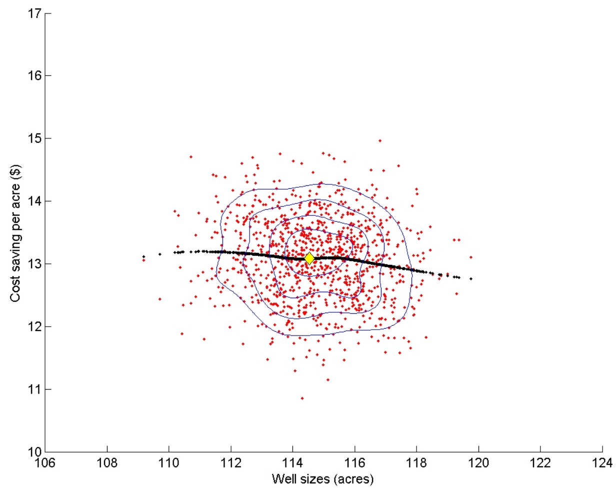
Each dot in this figure represents one draw of a 5 percent random sampling on well-base. The curve parallel to the x-axis is locally weighted scatterplot smoothing (Lowess) with bandwidth 0.9. The contours represent 20th, 40th, 60th, and 80th percentiles of a two dimensional kernel density. For example, the 20th percentile contour contains 200 draws out of 1000, and the range between 20th and 40th percentile also contains 200 draws, and so forth. The diamond shows the population mean of cost saving per acre and population mean of the well size.

Figure 4: Systematic Samples for Well-based Sampling



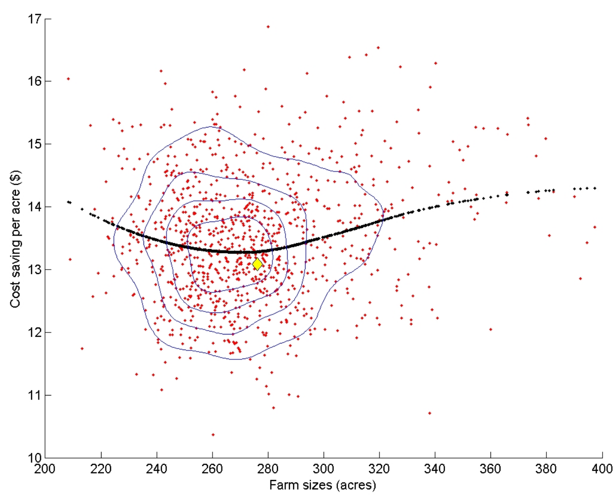
In 5 percent well-base systematic samplings, there are only 28 possible samples, since there are 10908 wells in total and the starting point lies between 1 and 28.

Figure 5: Stratified Samples for Well-based Sampling



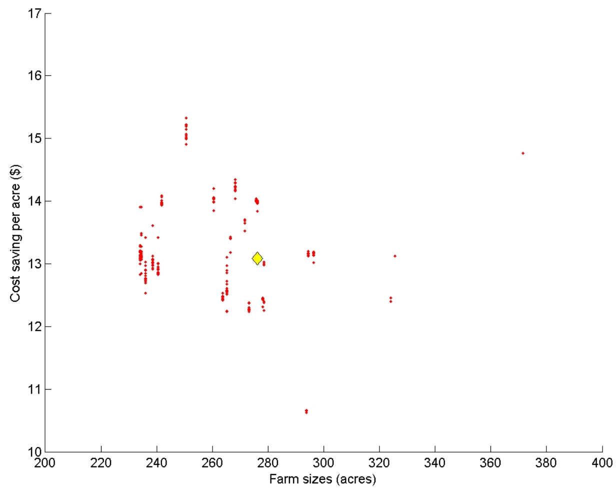
Refer to the footnote in Figure 3

Figure 6: Random Samples for Farm-based Sampling



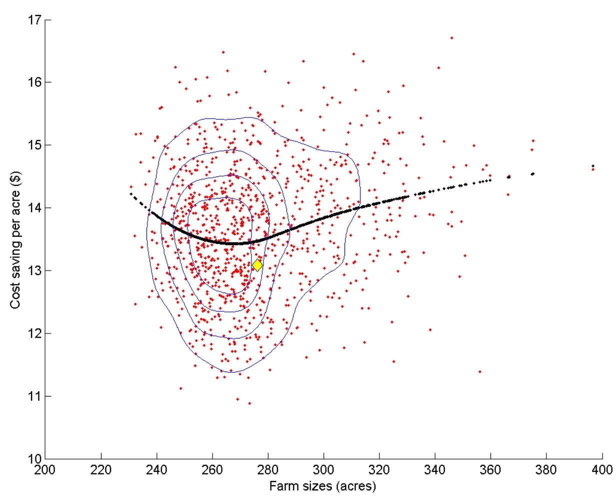
Each dot in this figure represents one draw of a 5 percent random sampling on farm-base. The definitions of Lowess smoother, contours and diamond refer to the footnote in Figure 3

Figure 7: Systematic Samples for Farm-based Sampling



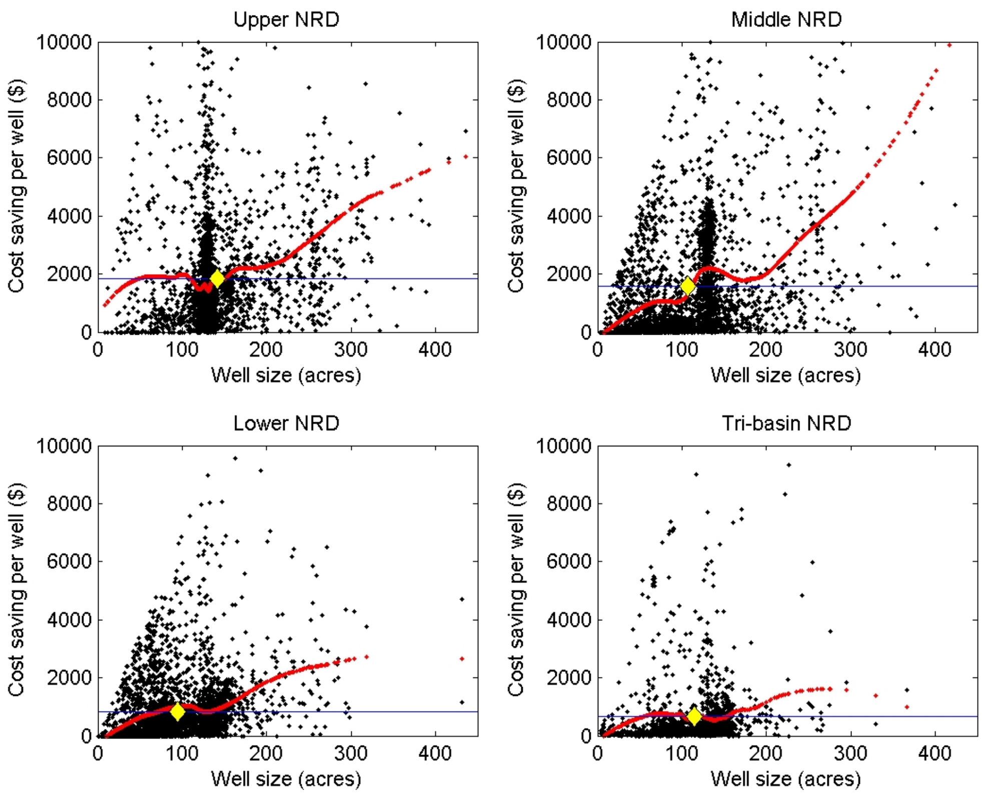
In 5 percent farm-base systematic samplings, there are only 25 possible samples, since there are 4525 wells in total and the starting point lies between 1 and 25.

Figure 8: Stratified Samples for Farm-based Sampling



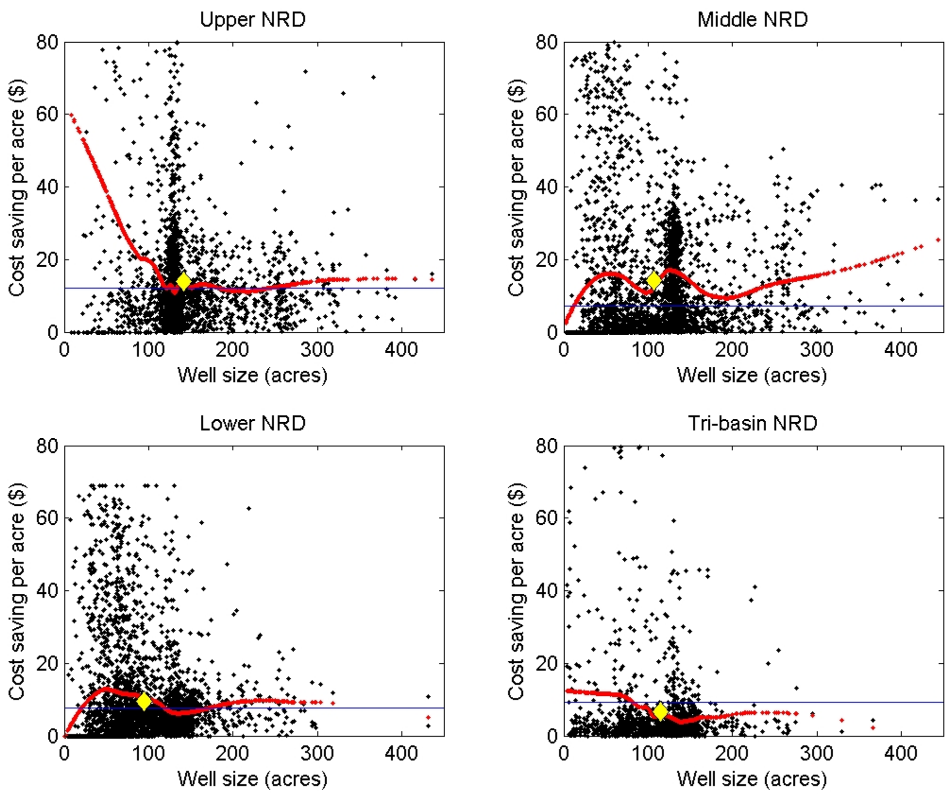
Refer to the footnote in Figure 6

Figure 9: Cost Saving Per Well in Four NRDs Using Well-based Sampling



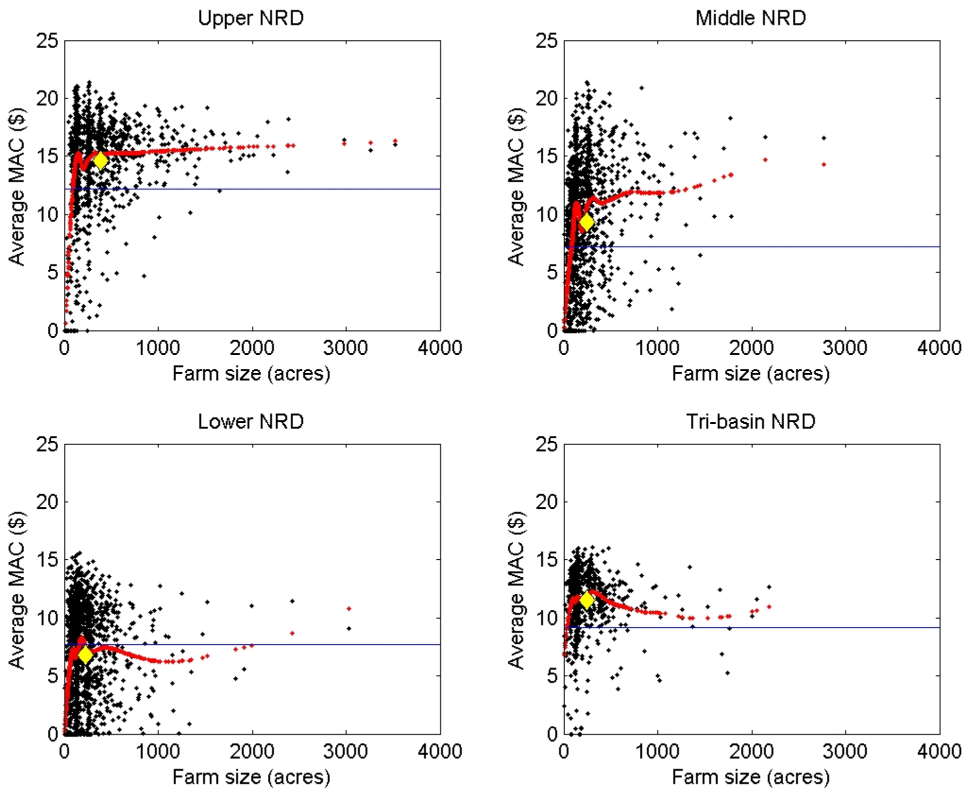
In each panel, the dots are population data for cost saving per well against well sizes in NRD-wide trading. The horizontal line is the equilibrium permit price in NRD-wide trading. The diamond represents the population means of cost savings per well and well sizes

Figure 10: Cost Saving Per Acre in Four NRDs Using Well-based Sampling



In each panel, the dots are population data for cost saving per acre of a well against its well size.
The horizontal line is the equilibrium permit price in NRD-wide trading. The diamond represents the population means of cost savings per acre and well sizes

Figure 11: Cost Saving Per Acre in Four NRDs Using Farm-based Sampling



In each panel, the dots are population data for average marginal abatement costs weighted by area of a farm against its farm size.
This figure only shows the farms smaller than 4000 acres. Figure A-6 shows all the farms.
The horizontal lines are equilibrium permit prices in NRD-wide trading. The diamond represents the population means of average marginal abatement costs weighted by area and farm sizes

Appendix A: Supplemental Tables and Figures

Table A-1: Certification and Trading in NRDs

NRD	Certification of Irrigated Acreage	Metering of Well	Current Allocation	Transfer Requirements
Upper Republican	1977	1982	13	Within township
Middle Republican	2003	2004	12	Within "sub-area"
Lower Republican	2004	2004	9	No transfers allowed
Tri-Basin	2004	2003	9-11	Permit application for >1 mile

Reproduced from Palazzo (2009).
The first and second columns are the starting time for certification and well metering.
The third column is current allocation measured by inches per acre. In Tri-basin NRD, the allocation is 9 in Kearney County, 10 in Phelps County, and 11 in Gosper County.(Nebraska Department of Natural Resources and Tri-basin Natural Resource District 2007)
The last column shows the transfer types in January 2008- November 2008. MRNRD Integrated Management Plan: Within "Quick response" sub-area, within "Upland" sub-area, or from "Quick response" sub-area to "Upland sub-area" (Nebraska Department of Natural Resources and Middle Republican Natural Resource District 2008).
URNRD Integrated Management Plan: "Floating township" described as a set of 36 quarter sections lying in a contiguous block; 6 blocks east to west and 6 blocks north to south (Nebraska Department of Natural Resources and Upper Republican Natural Resource District 2008).

Table A-2: Summary Statistics of Strata

		Number of Well	Mean Acreage Per Well	Number of Farm	Mean Acreage Per Farm	Total Acreage	% of Total Acreage
Small	[1, 160]	2742	90	2321	107	247533	19.81%
Medium	(160,320]	2459	112	1148	239	274293	21.96%
Large	(320, 14700]	5707	127	1056	689	727497	58.23%
Total		10908	115	4525	276	1249322	100.00%

Table A-3: Well-based Random Sampling on Basin-wide Trading

		Cost Saving (10^7) Scaled by the Number of Wells	Cost Saving (10^7) Scaled by Area	Permit Price
Population	Value	1.63	1.63	9.19
1% Sample	Mean	-1.21%	-1.25%	0.26%
	Median	-1.88%	-2.15%	0.25%
	std.	12.27%	11.49%	7.24%
2% Sample	Mean	-0.72%	-0.67%	0.19%
	Median	-1.22%	-0.64%	0.01%
	std.	8.42%	7.91%	5.26%
3% Sample	Mean	-0.56%	-0.44%	0.10%
	Median	-0.67%	-0.80%	0.21%
	std.	7.19%	6.76%	4.46%
4% Sample	Mean	-0.29%	-0.24%	0.06%
	Median	-0.42%	-0.32%	0.04%
	std.	5.99%	5.66%	3.84%
5% Sample	Mean	-0.21%	-0.19%	0.03%
	Median	-0.07%	-0.03%	-0.06%
	std.	5.37%	5.06%	3.49%

The first row shows the population values in basin-wide trading. The following rows present the percentage of biases or standard deviations in 1000 draws.

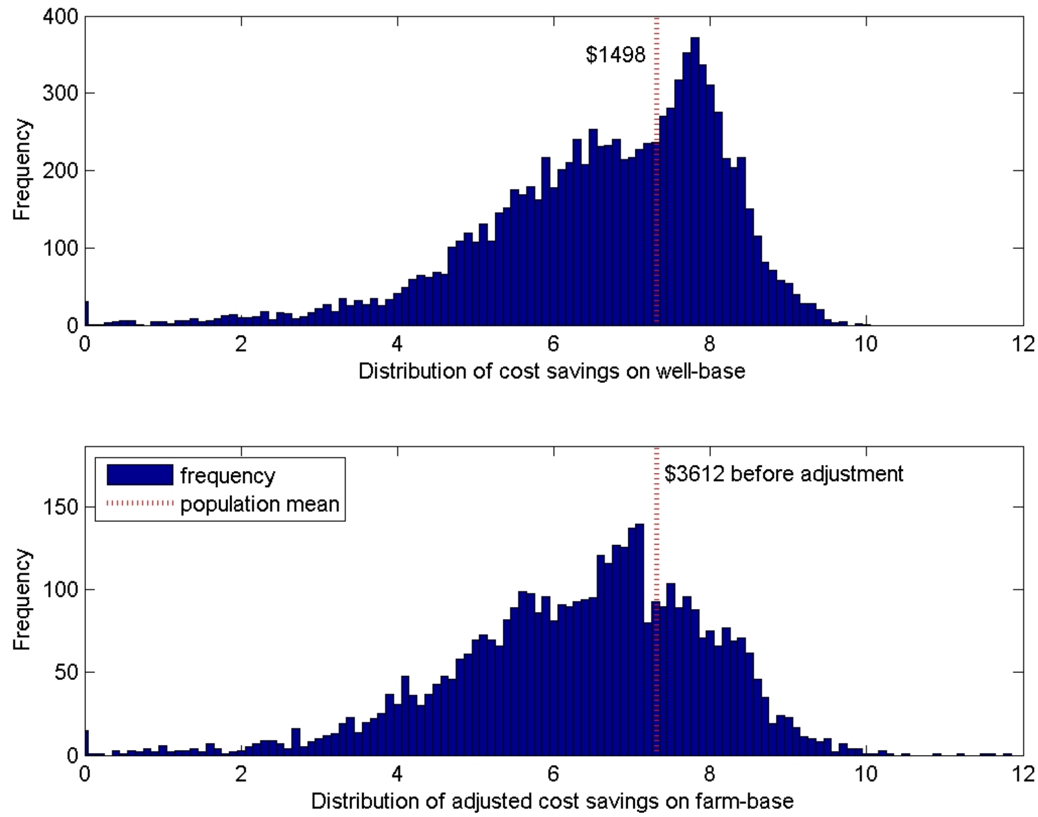
Table A-4: Farm-based Random Sampling on Basin-wide Trading

		Cost Saving (10^7) Scaled by the Number of Wells	Cost Saving (10^7) Scaled by Area	Permit Price
Population	Value	1.63	1.63	9.19
1% Sample	Mean	32.62%	35.22%	32.33%
	Median	28.20%	29.22%	31.32%
	std.	25.33%	26.29%	10.12%
2% Sample	Mean	8.26%	9.01%	18.74%
	Median	7.33%	8.09%	18.56%
	std.	13.39%	12.43%	6.38%
3% Sample	Mean	3.54%	3.91%	13.76%
	Median	2.82%	3.59%	13.90%
	std.	10.99%	9.80%	5.50%
4% Sample	Mean	3.57%	3.86%	13.53%
	Median	3.08%	3.42%	13.92%
	std.	9.43%	8.37%	4.56%
5% Sample	Mean	2.78%	3.00%	12.19%
	Median	1.88%	2.62%	12.61%
	std.	8.59%	7.41%	4.03%

Table A-5: How Much the Farm-based Sampling Enlarges the Biases

	Random	Systematic	Stratified
	Bias in Cost Savings		
Well-base	-0.21%	-0.36%	0.03%
Farm-base	2.78%	1.30%	7.13%
Farm-base Bias/Well-base Bias	-13	-4	238
	Bias in Permit Prices		
Well-base	0.03%	-0.41%	0.15%
Farm-base	12.19%	9.62%	13.09%
Farm-base Bias/Well-base Bias	448	-23	87

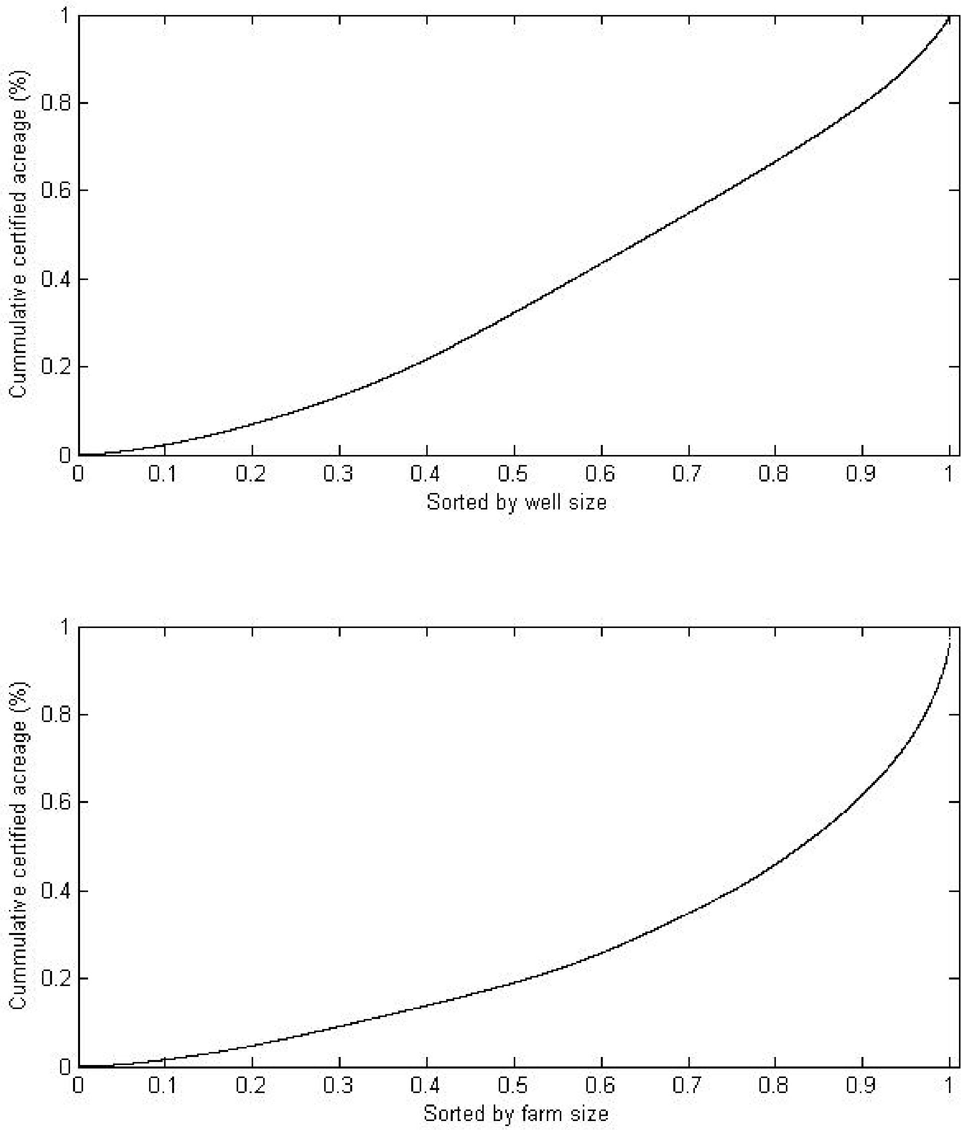
Figure A-1: Distribution of Estimated Cost Savings, Well-based and Farm-based



The upper panel shows the distribution of estimated cost savings on well-base, using the population data. The x-axis is logarithm of cost savings for each well, while the y-axis shows the frequency of wells at each value. The dotted line is population mean of cost saving per well.

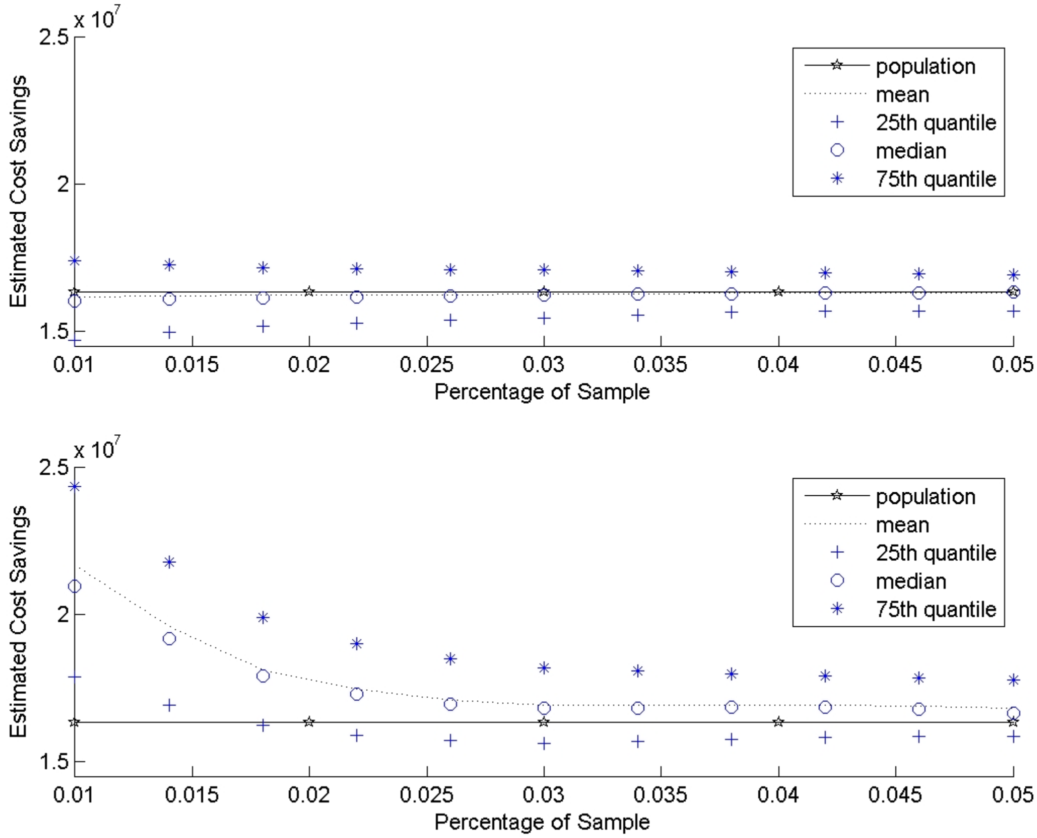
The lower panel shows the distribution of estimated cost savings on farm-base, using the population data. The x-axis is logarithm of cost savings for each farm, while the y-axis shows the frequency of farms at each value. To make the well-base cost savings comparable to those on farm-base in terms of frequency, I lengthened the y-axis in the lower panel by (number of wells)/(number of farms). Horizontally, the farm-base cost savings is divided by (number of wells)/(number of farms) before logarithmized, since one farm owns (number of wells)/(number of farms) wells on average. The dotted line is population mean of cost saving per farm.

Figure A-2: Cumulative Distribution of Certified Acreage



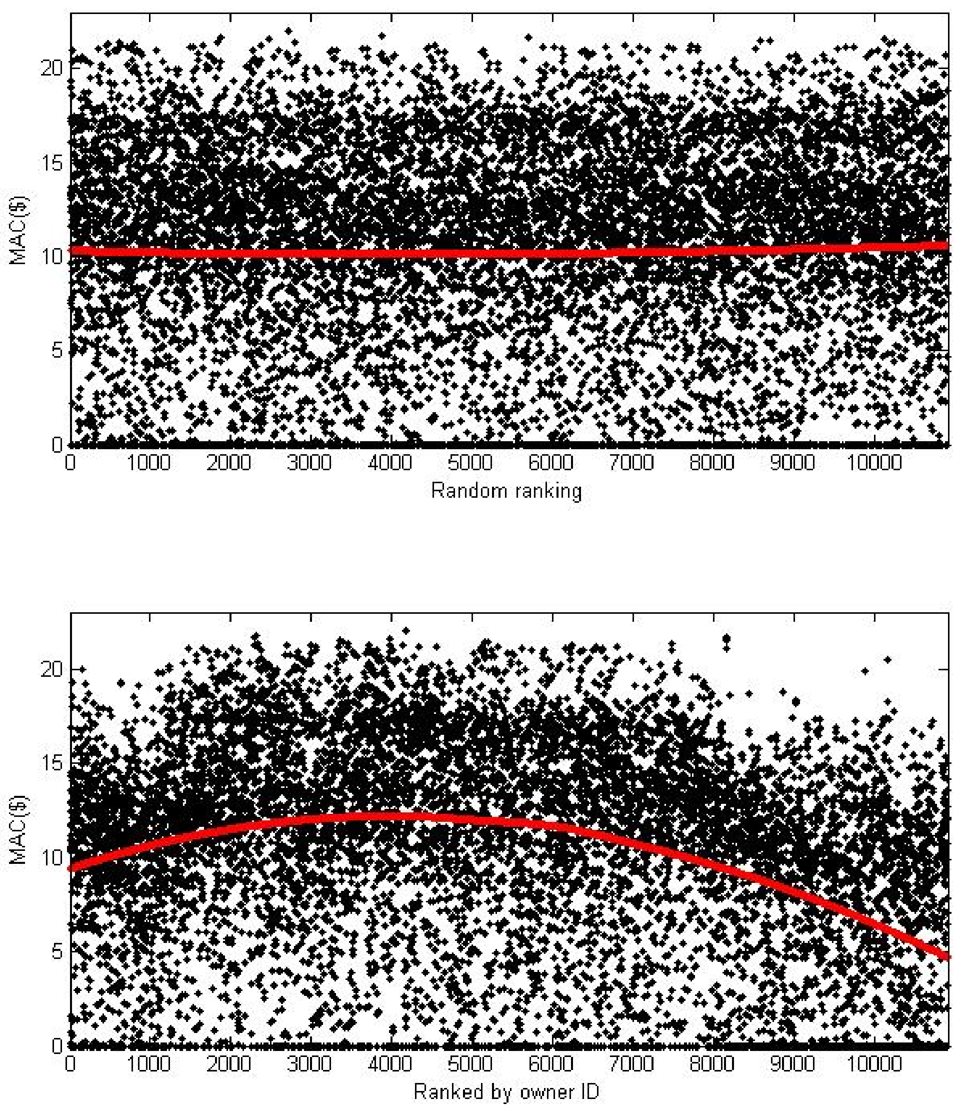
The upper panel shows the cumulative distribution of the certified acreage on well-based data. The lower panel shows the cumulative distribution of the certified acreage on farm-based data. The curves in both panels depart from the diagonal lines, which implies the existence of some extremely large values. In the lower panel, the sharp increase at the right end means that the 10 percent largest farms account for almost 40 percent of the total certified acreage.

Figure A-3: The Trend in Estimates as the Sample Size Rises



As the sample sizes in simple random sampling increase from 1 percent to 5 percent in Table A-3 and Table A-4 in the Appendix, we can see a significant reduction in the means and standard deviations of biases.
In the upper panel, the biases in well-base samplings are below the population value, but very close to it.
In the lower panel, the biases in farm-base samplings start about 1/3 higher than the population value, and then move closely to the population value after 3 percent. The distances between 25th quantile and 75th quantile are larger than those in the upper panel.

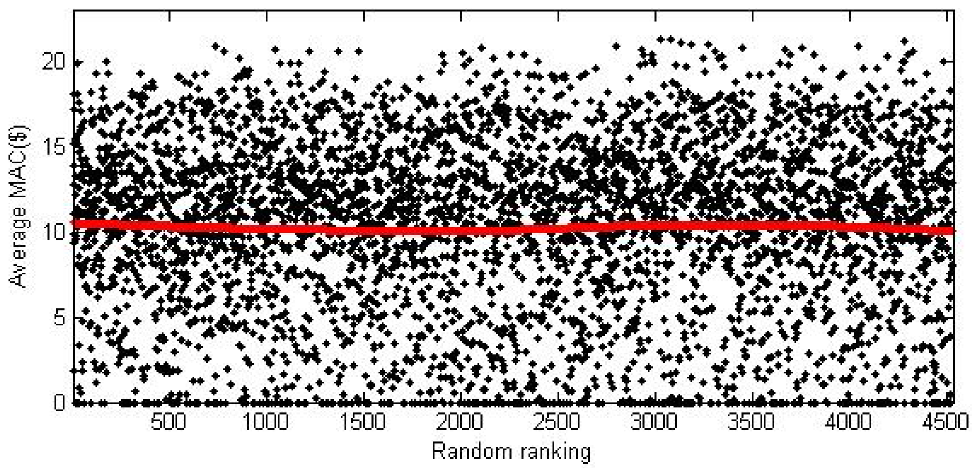
Figure A-4: Marginal Abatement Costs Ranked for Well-based Sampling

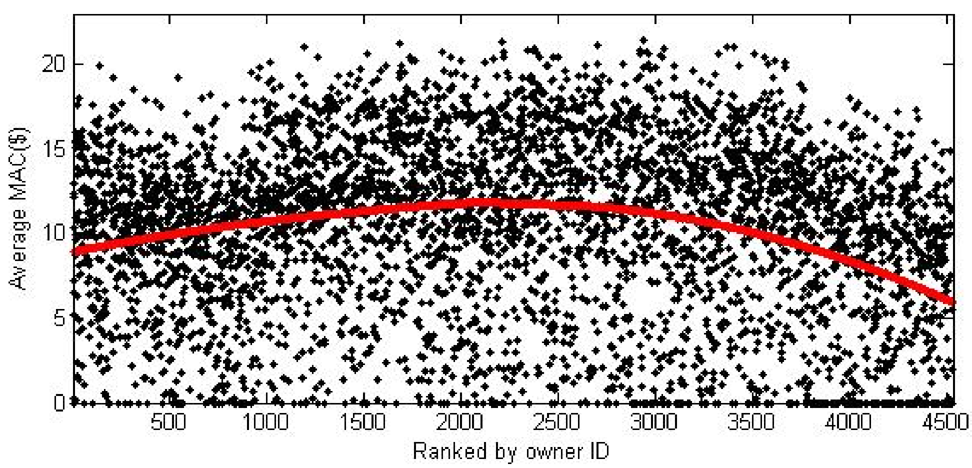


In the upper panel, the marginal abatement costs for all wells in population are randomly ranked. There does not exist an obvious trend.

In the lower panel, the marginal abatement costs for all wells in population are ranked by owner IDs, which are issued depending on the order of well registration.The best fields are registered earlier, so the trend of marginal abatement costs goes up at the beginning. At around the middle point, average marginal abatement cost starts to drop down, because most of registered wells since then are used to irrigate low quality land, where the marginal benefit from irrigation is very small.

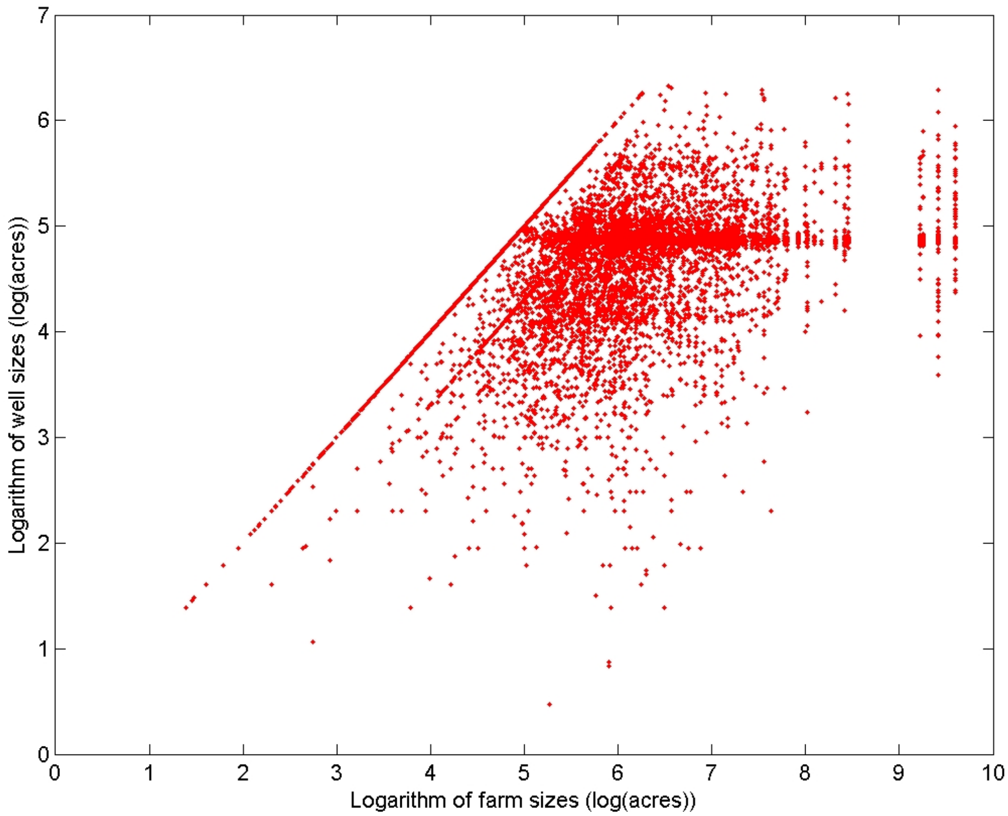
Figure A-5: Marginal Abatement Costs Ranked for Farm-based Sampling





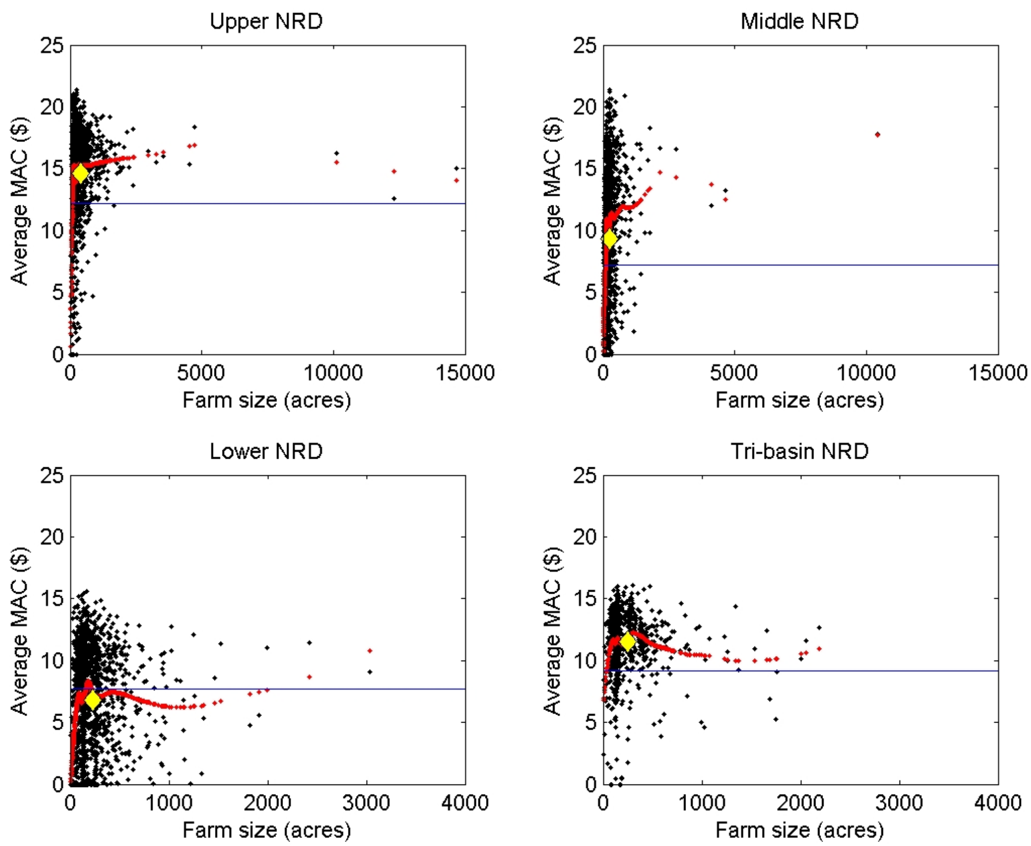
Sampling on farm-base.
Refer to footnotes in Figure A-4.

Figure A-6: Well Sizes Against Farm Sizes



This figure shows the logarithm of the well sizes against the logarithm of the farm sizes. The diagonal line represent those farms which have only one well. The pattern in this figure implies most wells in large farms are also large wells.

Figure A-7: Cost Saving Per Well in Four NRDs for Farm-based Sampling



In each panel, the dots are population data for average marginal abatement costs weighted by area of a farm against its farm size.
The horizontal lines are equilibrium permit prices in NRD-wide trading.

Appendix B: Computer Code

Well-base Random Sampling

In each draw, generate the sample by random number first.Then, use this code to generate the optimal permit prices, irrigation water allocation, and calculate the cost saving in abatement. Record these information for each single draw and save these data for further analysis. This is an example for 5 percent well-base random sampling. I also replaced the ratio in the 11^{th} line in this code with 0.01 to 0.04 to get 1 percent to 4 percent samples.

```
1  % First set for Mac/R2008a
2  %options = optimset('LargeScale','off','GradObj','on','TolFun',1e-8,'Display','off',...
3  %'Algorithm','active-set');
4  options = optimset('LargeScale','off','GradObj','on','TolFun',1e-8,'Display','off');
5
6  %read in data
7  load MAC_3_24_09
8
9  % columns in wells
10 % 1     well_id
11 % 2     cert_acres
12 % 3     volume_acre_feet
13 % 4     irr_type
14 % 5     nrd
15 % 6     latdd
16 % 7     longdd
17 % 8     section_
18 % 9     township
19 % 10    rangenum
```

```
% 11    soil_type
% 12    regnum
% 13    replacemen
% 14    status
% 15    useid
% 16    nrdname
% 17    nrddwrnum
% 18    countyname
% 19    countynum
% 20    acres
% 21    gpm
% 22    ownernumbe
% 23    compname
% 24    city
% 25    st
% 26    zip
% 27    distance_to_nearest_stream
% 28    distance_to_nearest_well
% 29    well_id_nearest_well
% 30    cmpldyear
% 31     et_alfalfa
% 32     et_corn
% 33     et_beans
% 34     et_sorghum
% 35     et_soybeans
% 36     et_beets
% 37     et_wheat
% 38     yd_ym_alfalfa
% 39     yd_ym_corn
% 40     yd_ym_beans
% 41     yd_ym_sorghum
% 42     yd_ym_soybeans
% 43     yd_ym_beets
% 44     yd_ym_wheat
% 45     yd_default_corn
% 46     yd_default_sorghum
% 47     yd_default_soybeans
% 48     yd_default_wheat
% 49     ym_default_corn
% 50     ym_default_sorghum
% 51     ym_default_soybeans
% 52     ym_default_wheat
```

```
% 53     pumping water level
% 54     current NRD allocation
% 55       total acreage of the farm this well belongs to
% 56     how many wells owned by the farm this well belongs to
% 57       1:10908
% 58       is this well the first well of the farm it belongs to?


tic
% To get 5 percent samples
ratio=0.05;

% Take 1000 draws.
draws=1000;
num_sample=round(num_wells*ratio);

%set seed at 1
rand('seed',1);

for i=1:draws
    clear sample

r=[round(num_wells.*rand(num_wells,1)) linspace(1,num_wells,num_wells)'];
ordered_sample=sortrows(r);

% The sample for the ith draw
sample=sort(ordered_sample(1:num_sample,2));

five_value=value(sample,:);
five_inches=inches(sample,:);
five_acres_irrg=acres(sample,:);
five_acres=acres(sample,:) + dryland(sample,:);
five_well_id=well_id(sample,:);
five_dist_stream=dist_stream(sample,:);
five_dist_wells=dist_well(sample,:);
five_yield=yield(sample,:);
five_lift=lift(sample,:);
five_soil=soil(sample,:);
five_cert_acres= cert_acres(sample);
five_wells= wells{1}(sample);
five_pump_yield=pump_yield(sample);
five_soil_types=soil_types(sample);
```

```
% To trace out the MAC curve
for j=1:size(five_wells,1)
  p2.five(j,:)=p2.basin(sample(j),:);
  q2.five(j,:)=q2.basin(sample(j),:);
% Abatement (acre-inches) under current allocation (before trading)
initial_abate.five(j)= max(five_inches(j,39)-current_allocation(sample(j)),0)*five_acres(j,39);
% Abatement (inches per acre) under current allocation (before trading)
initial_abate_inches.five(j)= max(five_inches(j,39)-current_allocation(sample(j)),0);
temp_wells.five = size(five_wells,1);
current_use.five(j)=min(current_allocation(sample(j)),five_inches(j,39));
% The amount can be sold.
free_ride.five(j)=current_allocation(sample(j))-current_use.five(j);
free_ride.five(find(five_inches(j,39)==0))=0;
five_acres_i(j)=five_acres(j,39);
end
% Deal with the corner solution
for j=1:size(five_wells,1)
  if   p2.five(j,40)>0
       loc.start=find(p2.five(j,:)>0.0001,1,'first')-1;
%last zero value
       loc.end=(find(p2.five(j,:)>0.0001 & gradient(p2.five(j,:))<0.1,1,'first')-1);
%first zero gradient
       q0=q2.five(j,loc.start);%find baseline for normalization
       q_temp=q2.five(j,loc.start:loc.end);%pull out relevant points
       q_temp=[q_temp-ones(size(q_temp)).*q0 five_inches(j,39) five_inches(j,39)+.00001];
```

```
%normalize to origin
       q_temp=q_temp+ones(size(q_temp)).*free_ride.five(j);

%adjust for slack constraints and initial water use

       p_temp=[p2.five(j,loc.start:loc.end) p2.five(j,loc.end) 100];

q_junk.five(j,:)=linspace(min(q_temp),max(q_temp)- 0.01001,48);

p_i.five(j,:)=[pchip(q_temp,p_temp,q_junk.five(j,:)) p2.five(j,loc.end) 100];

q_i.five(j,:)=[q_junk.five(j,:) max(five_inches(j,39),current_allocation(sample(j))) max(five_inches
(j,39),current_allocation(sample(j)))+0.0001];

p_i.five(j,:) = p_i.five(j,:) + linspace(0,1e-6,size(p_i.five(j,:),2));

  elseif p2.five(j,40)<=0

      p_i.five(j,:)= zeros(1,50);
      p_i.five(j,1)= 1e10;

% need to give q_i a number to avoid breaking of the loop.
      q_i.five(j,:)= zeros(1,50);

  end

end


%find the mac (tax) where inches abated equals abatement under NRD allocations
tax.five(i) = fzero(@(u) sum(pigou(u,temp_wells.five,p_i.five,q_i.five,five_acres_i))
-sum(initial_abate.five),10);

%vector of macs
mac.tax.five = tax.five(i)*ones(size(five_wells,1),1);


for j=1:size(five_wells,1)

    if p_i.five(j,50)>0

 reduced_allocation.five(j) = pchip(p_i.five(j,:),q_i.five(j,:),mac.tax.five(j))*five_acres(j,39);
```

```
%find acre inches reduced under equalized mac

 reduced_allocation_inches.five(j)=pchip(p_i.five(j,:),q_i.five(j,:),mac.tax.five(j));

%find inches per acre reduced under equalized mac

    elseif p_i.five(j,50)≤0

 reduced_allocation.five(j) =0; %for dryland wells

 reduced_allocation_inches.five(j)= 0; %for dryland wells

    end

acreinches_traded.five(j)=reduced_allocation.five(j)−initial_abate.five(j);
temp_free.five(j)=free_ride.five(j)*five_acres(j,39);
inches_traded.five(j)=reduced_allocation_inches.five(j)−initial_abate_inches.five(j);

end

for j=1:size(five_wells,1)

 if q_i.five(j,50)>0
q_tc_interp.five(j,:)=linspace(min(q_i.five(j,:)),max(q_i.five(j,:)),101);

tc_interp.five(j,:) = pchip(q_i.five(j,:),p_i.five(j,:),q_tc_interp.five(j,:))
*triu(ones(101))/100*(max(q_i.five

(j,:))−min(q_i.five(j,:)));

    if q_tc_interp.five(j,1)>0

        tc.abate.five(j)=0;
        mac.abate.five(j)=0;

    else
       tc.abate.five(j)=pchip(q_tc_interp.five(j,:),...
           [tc_interp.five(j,1:find(q_tc_interp.five(j,:)<max(q_tc_interp.five(j,:)),1,'last'))...
           tc_interp.five(j,find(q_tc_interp.five(j,:)<max(q_tc_interp.five(j,:)),1,'last'))

+.00001],initial_abate_inches.five(j));

```

```
        mac.abate.five(j)=pchip(q_i.five(j,:),p_i.five(j,:),initial_abate_inches.five(j));

    end


tc.trade.five(j)=pchip(q_tc_interp.five(j,:),...
    [tc_interp.five(j,1:find(q_tc_interp.five(j,:)<max(q_tc_interp.five(j,:)),1,'last'))...
    tc_interp.five(j,find(q_tc_interp.five(j,:)<max(q_tc_interp.five(j,:)),1,'last'))...
        +.00001],reduced_allocation_inches.five(j));

 else
     tc.abate.five(j)=0;
     tc.trade.five(j)=0;

 end

% Get the cost saving from trading
market_cost_trade.five(j)= inches_traded.five(j)*mac.tax.five(j);

% cost saving for each well in this sample in the ith draw.
welfare_gain.five(j) = (tc.abate.five(j)- tc.trade.five(j)
+ market_cost_trade.five(j))*five_acres(j,39);

if welfare_gain.five(j)<0
    welfare_gain.five(j)=0;
end

end

% Scale the sample estimates back to population estimates
sum_gain.five(i)=sum(welfare_gain.five)*num_wells/num_sample;
tot_area.five(i)=sum(wells{2}(sample));
sum_gain_area.five(i)=sum(welfare_gain.five)*sum(wells{2})/tot_area.five(i);

% Get other characteristics of the samples
tot_lift.well_five(i)=mean(lift(sample));
tot_yield.well_five(i)=mean(yield(sample));

% The number of buyers, sellers and outsiders.
wbuy.erfive(i)=sum(market_cost_trade.five<0);
wout.erfive(i)=sum(market_cost_trade.five==0);
wsell.erfive(i)=sum(market_cost_trade.five>0);
```

```

% The percentage of buyers, sellers and outsiders.
wbuy.erfive_r(i)=wbuy.erfive(i)/size(sample,1);
wout.erfive_r(i)=wout.erfive(i)/size(sample,1);
wsell.erfive_r(i)=wsell.erfive(i)/size(sample,1);

% People who sell all their permits and move to dryland
wsell.all_five(i)=sum(inches_traded.five'≥current_allocation(sample))/size(sample,1);

% Amount of permits traded
wsell.traded_five(i)=sum(abs(inches_traded.five)*five_acres(:,39))/2;

end
toc
% save it
```

Well-base Systematic Sampling

```
tic
% To get 5 percent samples
ratio=0.05;

% Take 1000 draws.
draws=1000;
num_sample=round(num_wells*ratio);
r= sortrows([linspace(1,num_wells,num_wells)' wells{2}],2);

%set seed at 1
rand('seed',1);

% Random starting point
r_start=ceil(rand(draws,1)*(num_wells-1/ratio*(num_sample-1)));

for i=1:draws
 clear sample

% From a random start, select every 20th well.
sample_r=linspace(r_start(i),r_start(i)+1/ratio*(num_sample-1),num_sample)';

% The sample for the ith draw
sample=r(sample_r);

sys5_value=value(sample,:);
sys5_inches=inches(sample,:);
sys5_acres_irrg=acres(sample,:);
sys5_acres=acres(sample,:) + dryland(sample,:);
sys5_well_id=well_id(sample,:);
sys5_dist_stream=dist_stream(sample,:);
sys5_dist_wells=dist_well(sample,:);
sys5_yield=yield(sample,:);
sys5_lift=lift(sample,:);
sys5_soil=soil(sample,:);
sys5_cert_acres= cert_acres(sample);
sys5_wells= wells{1}(sample);
sys5_pump_yield=pump_yield(sample);
```

```
sys5_soil_types=soil_types(sample);

for j=1:size(sys5_wells,1)

  p2.sys5(j,:)=p2.basin(sample(j),:);
  q2.sys5(j,:)=q2.basin(sample(j),:);

initial_abate.sys5(j)= max(sys5_inches(j,39)-current_allocation(sample(j)),0)*sys5_acres(j,39);

initial_abate_inches.sys5(j)= max(sys5_inches(j,39)-current_allocation(sample(j)),0);


temp_wells.sys5 = size(sys5_wells,1);

current_use.sys5(j)=min(current_allocation(sample(j)),sys5_inches(j,39));
free_ride.sys5(j)=current_allocation(sample(j))-current_use.sys5(j);
free_ride.sys5(find(sys5_inches(j,39)==0))=0;
sys5_acres_i(j)=sys5_acres(j,39);

end

for j=1:size(sys5_wells,1)

  if   p2.sys5(j,40)>0

       loc.start=find(p2.sys5(j,:)>0.0001,1,'first')-1;%last zero value
      %first zero gradient
       loc.end=(find(p2.sys5(j,:)>0.0001 & gradient(p2.sys5(j,:))<0.1,1,'first')-1);

       q0=q2.sys5(j,loc.start);%find baseline for normalization

       q_temp=q2.sys5(j,loc.start:loc.end);%pull out relevant points
       %normalize to origin
       q_temp=[q_temp-ones(size(q_temp)).*q0 sys5_inches(j,39) sys5_inches(j,39)+.00001];
      %adjust for slack constraints and initial water use
       q_temp=q_temp+ones(size(q_temp)).*free_ride.sys5(j);

       p_temp=[p2.sys5(j,loc.start:loc.end) p2.sys5(j,loc.end) 100];

       %the problem is below
q_junk.sys5(j,:)=linspace(min(q_temp),max(q_temp)- 0.01001,48);
```

```
p_i.sys5(j,:)=[pchip(q_temp,p_temp,q_junk.sys5(j,:)) p2.sys5(j,loc.end) 100];

q_i.sys5(j,:)=[q_junk.sys5(j,:) max(sys5_inches(j,39),current_allocation(sample(j)))
max(sys5_inches(j,39),current_allocation(sample(j)))+0.0001];

p_i.sys5(j,:) = p_i.sys5(j,:) + linspace(0,1e-6,size(p_i.sys5(j,:),2));

  elseif p2.sys5(j,40)≤0

      p_i.sys5(j,:)= zeros(1,50);
      p_i.sys5(j,1)= 1e10;
      q_i.sys5(j,:)= zeros(1,50);

  end

end


%find the mac (tax) where inches abated equals abatement under NRD allocations
tax.sys5(i) = fzero(@(u) sum(pigou(u,temp_wells.sys5,p_i.sys5,q_i.sys5,sys5_acres_i))
-sum(initial_abate.sys5),10);

mac.tax.sys5 = tax.sys5(i)*ones(size(sys5_wells,1),1); %vector of macs

for j=1:size(sys5_wells,1)

    if p_i.sys5(j,50)>0
 %find acre inches reduced under equalized mac
 reduced_allocation.sys5(j) = pchip(p_i.sys5(j,:),q_i.sys5(j,:),mac.tax.sys5(j))*sys5_acres(j,39);
 %find inches per acre reduced under equalized mac
 reduced_allocation_inches.sys5(j)=pchip(p_i.sys5(j,:),q_i.sys5(j,:),mac.tax.sys5(j));

    elseif p_i.sys5(j,50)≤0

 reduced_allocation.sys5(j) =0; %for dryland wells

 reduced_allocation_inches.sys5(j)= 0; %for dryland wells

    end

acreinches_traded.sys5(j)=reduced_allocation.sys5(j)-initial_abate.sys5(j);

```

```
temp_free.sys5(j)=free_ride.sys5(j)*sys5_acres(j,39);

inches_traded.sys5(j)=reduced_allocation_inches.sys5(j)-initial_abate_inches.sys5(j);

end


for j=1:size(sys5_wells,1)

 if q_i.sys5(j,50)>0
q_tc_interp.sys5(j,:)=linspace(min(q_i.sys5(j,:)),max(q_i.sys5(j,:)),101);

tc_interp.sys5(j,:) = pchip(q_i.sys5(j,:),p_i.sys5(j,:),q_tc_interp.sys5(j,:))
*triu(ones(101))/100*(max(q_i.sys5(j,:))-min(q_i.sys5(j,:)));


    if q_tc_interp.sys5(j,1)>0

        tc.abate.sys5(j)=0;
        mac.abate.sys5(j)=0;

    else
       tc.abate.sys5(j)=pchip(q_tc_interp.sys5(j,:),...
           [tc_interp.sys5(j,1:find(q_tc_interp.sys5(j,:)<max(q_tc_interp.sys5(j,:)),1,'last'))...
           tc_interp.sys5(j,find(q_tc_interp.sys5(j,:)<max(q_tc_interp.sys5(j,:)),1,'last'))
           +.00001],initial_abate_inches.sys5(j));

       mac.abate.sys5(j)=pchip(q_i.sys5(j,:),p_i.sys5(j,:),initial_abate_inches.sys5(j));

    end

tc.trade.sys5(j)=pchip(q_tc_interp.sys5(j,:),...
    [tc_interp.sys5(j,1:find(q_tc_interp.sys5(j,:)<max(q_tc_interp.sys5(j,:)),1,'last'))...
    tc_interp.sys5(j,find(q_tc_interp.sys5(j,:)<max(q_tc_interp.sys5(j,:)),1,'last'))+.00001],
    reduced_allocation_inches.sys5(j));

 else
     tc.abate.sys5(j)=0;
     tc.trade.sys5(j)=0;

 end

```

```
market_cost_trade.sys5(j)= inches_traded.sys5(j)*mac.tax.sys5(j);

welfare_gain.sys5(j) = (tc.abate.sys5(j)- tc.trade.sys5(j)
+ market_cost_trade.sys5(j))*sys5_acres(j,39);

if welfare_gain.sys5(j)<0
    welfare_gain.sys5(j)=0;
end

end

% Scale the sample estimates back to population estimates
sum_gain.sys5(i)=sum(welfare_gain.sys5)*num_wells/num_sample;
tot_area.sys5(i)=sum(wells{2}(sample));
sum_gain_area.sys5(i)=sum(welfare_gain.sys5)*sum(wells{2})/tot_area.sys5(i);

% Get other characteristics of the samples
tot_lift.well_sys5(i)=mean(lift(sample));
tot_yield.well_sys5(i)=mean(yield(sample));

% The number of buyers, sellers and outsiders.
wbuy.ersys5(i)=sum(market_cost_trade.sys5<0);
wout.ersys5(i)=sum(market_cost_trade.sys5==0);
wsell.ersys5(i)=sum(market_cost_trade.sys5>0);

% The percentage of buyers, sellers and outsiders.
wbuy.ersys5_r(i)=wbuy.ersys5(i)/size(sample,1);
wout.ersys5_r(i)=wout.ersys5(i)/size(sample,1);
wsell.ersys5_r(i)=wsell.ersys5(i)/size(sample,1);

% People who sell all their permits and move to dryland
wsell.all_sys5(i)=sum(inches_traded.sys5'≥current_allocation(sample))/size(sample,1);

% Amount of permits traded
wsell.traded_sys5(i)=sum(abs(inches_traded.sys5)*sys5_acres(:,39))/2;

end
toc

% save it
```

Well-base Stratified Sampling

The strata used for this code are in Table A-2.

```
% Define small, medium and large wells.
small=(wells{2}≤160);
medium=(wells{2}>160 & wells{2}≤320);
large=(wells{2}>320);

% The number of wells in each stratum
num.small=sum(small);
num.medium=sum(medium);
num.large=sum(large);

tic
% To get 5 percent samples
ratio=0.05;

% Take 1000 draws.
draws=1000;

num_sample=round(num_wells*ratio);
num_sample_small=round(num.small*ratio);
num_sample_medium=round(num.medium*ratio);
num_sample_large=round(num.large*ratio);

%set seed at 1
rand('seed',1);


for i=1:draws
    clear sample_f
r=rand(num_wells,1);

% Get random wells from each stratum.
r_s=[round(num_wells.*r) linspace(1,num_wells,num_wells)'.*small wells{1}];
ordered_sample_s=sortrows(r_s,-2);
ordered_sample_ss=sortrows(ordered_sample_s(1:sum(small),:));
sample_s=sort(ordered_sample_ss(1:num_sample_small,2));
```

```
tot_area.well_s(i)=sum(wells{2}(sample_s));

r_m=[round(num_wells.*r) linspace(1,num_wells,num_wells)'.*medium wells{1}];
ordered_sample_m=sortrows(r_m,-2);
ordered_sample_mm=sortrows(ordered_sample_m(1:sum(medium),:));
sample_m=sort(ordered_sample_mm(1:num_sample_medium,2));
tot_area.well_m(i)=sum(wells{2}(sample_m));


r_l=[round(num_wells.*r) linspace(1,num_wells,num_wells)'.*large wells{1}];
ordered_sample_l=sortrows(r_l,-2);
ordered_sample_ll=sortrows(ordered_sample_l(1:sum(large),:));
sample_l=sort(ordered_sample_ll(1:num_sample_large,2));
tot_area.well_l(i)=sum(wells{2}(sample_l));

% Combine wells from each stratum into a sample for ith draw.
sample_f=[sample_s; sample_m; sample_l];

five_STRAwell_value=value(sample_f,:);
five_STRAwell_inches=inches(sample_f,:);
five_STRAwell_acres_irrg=acres(sample_f,:);
five_STRAwell_acres=acres(sample_f,:) + dryland(sample_f,:);
five_STRAwell_well_id=well_id(sample_f,:);
five_STRAwell_dist_stream=dist_stream(sample_f,:);
five_STRAwell_dist_wells=dist_well(sample_f,:);
five_STRAwell_yield=yield(sample_f,:);
five_STRAwell_lift=lift(sample_f,:);
five_STRAwell_cert_acres= cert_acres(sample_f,:);
five_STRAwell_wells= wells{1}(sample_f,:);
tot_area.five_STRAwell(i)=sum(wells{2}(sample_f));

for j=1:size(sample_f,1)
p2.five_STRAwell(j,:)=wrev(a.p.basin(sample_f(j),:));

q2.five_STRAwell(j,:)=wrev(a.q.basin(sample_f(j),:));

initial_abate.five_STRAwell(j)= max(five_STRAwell_inches(j,39)-current_allocation(sample_f(j)),0)
*five_STRAwell_acres(j,39);

initial_abate_inches.five_STRAwell(j)= max(five_STRAwell_inches(j,39)
-current_allocation(sample_f(j)),0);
```

```
temp_wells.five_STRAwell = size(five_STRAwell_wells,1);

current_use.five_STRAwell(j)=min(current_allocation(sample_f(j)),five_STRAwell_inches(j,39));
free_ride.five_STRAwell(j)=current_allocation(sample_f(j))-current_use.five_STRAwell(j);
free_ride.five_STRAwell(find(five_STRAwell_inches(j,39)==0))=0;
five_STRAwell_acres_i(j)=five_STRAwell_acres(j,39);

end

for j=1:size(sample_f,1)

  if   p2.five_STRAwell(j,40)>0

        loc.start=find(p2.five_STRAwell(j,:)>0.0001,1,'first')-1;%last zero value

        loc.end=(find(p2.five_STRAwell(j,:)>0.0001 & gradient(p2.five_STRAwell(j,:))<0.1,
        1,'first')-1);%first zero gradient

        q0=q2.five_STRAwell(j,loc.start);%find baseline for normalization

        q_temp=q2.five_STRAwell(j,loc.start:loc.end);%pull out relevant points

        q_temp=[q_temp-ones(size(q_temp)).*q0 five_STRAwell_inches(j,39)
        five_STRAwell_inches(j,39)+.00001];%normalize to origin
       %adjust for slack constraints and initial water use
        q_temp=q_temp+ones(size(q_temp)).*free_ride.five_STRAwell(j);

        p_temp=[p2.five_STRAwell(j,loc.start:loc.end) p2.five_STRAwell(j,loc.end) 100];

        %the problem is below
q_junk.five_STRAwell(j,:)=linspace(min(q_temp),max(q_temp)- 0.01001,48);

p_i.five_STRAwell(j,:)=[pchip(q_temp,p_temp,q_junk.five_STRAwell(j,:))
p2.five_STRAwell(j,loc.end) 100];

q_i.five_STRAwell(j,:)=[q_junk.five_STRAwell(j,:) max(five_STRAwell_inches(j,39),
current_allocation(sample_f(j))) max(five_STRAwell_inches(j,39),
current_allocation(sample_f(j)))+0.0001];


p_i.five_STRAwell(j,:) = p_i.five_STRAwell(j,:) + linspace(0,1e-6,size(p_i.five_STRAwell(j,:),2));

```

```
  elseif p2.five_STRAwell(j,40)≤0

      p_i.five_STRAwell(j,:)= zeros(1,50);
      p_i.five_STRAwell(j,1)= 1e10;
      q_i.five_STRAwell(j,:)= zeros(1,50);

  end

end


%find the mac (tax) where inches abated equals abatement under NRD allocations
tax.five_STRAwell(i) = fzero(@(u) sum(pigou(u,temp_wells.five_STRAwell,p_i.five_STRAwell,
q_i.five_STRAwell,five_STRAwell_acres_i))-sum(initial_abate.five_STRAwell),10);

%vector of macs
mac.tax.five_STRAwell = tax.five_STRAwell(i)*ones(size(five_STRAwell_wells,1),1);

for j=1:size(five_STRAwell_wells,1)


    if p_i.five_STRAwell(j,50)>0

 %find acre inches reduced under equalized mac
 reduced_allocation.five_STRAwell(j) = pchip(p_i.five_STRAwell(j,:),q_i.five_STRAwell(j,:),
 mac.tax.five_STRAwell(j))*five_STRAwell_acres(j,39);
 %find inches per acre reduced under equalized mac
 reduced_allocation_inches.five_STRAwell(j)=pchip(p_i.five_STRAwell(j,:),q_i.five_STRAwell(j,:),
 mac.tax.five_STRAwell(j));

    elseif p_i.five_STRAwell(j,50)≤0

 reduced_allocation.five_STRAwell(j) =0; %for dryland wells

 reduced_allocation_inches.five_STRAwell(j)= 0; %for dryland wells

    end

acreinches_traded.five_STRAwell(j)=reduced_allocation.five_STRAwell(j)
-initial_abate.five_STRAwell(j);

temp_free.five_STRAwell(j)=free_ride.five_STRAwell(j)*five_STRAwell_acres(j,39);
```

```

inches_traded.five_STRAwell(j)=reduced_allocation_inches.five_STRAwell(j)
-initial_abate_inches.five_STRAwell(j);

end


for j=1:size(five_STRAwell_wells,1)

 if q_i.five_STRAwell(j,50)>0
q_tc_interp.five_STRAwell(j,:)=linspace(min(q_i.five_STRAwell(j,:)),max(q_i.five_STRAwell(j,:)),101);

tc_interp.five_STRAwell(j,:) = pchip(q_i.five_STRAwell(j,:),p_i.five_STRAwell(j,:),
q_tc_interp.five_STRAwell(j,:))
*triu(ones(101))/100*(max(q_i.five_STRAwell(j,:))-min(q_i.five_STRAwell(j,:)));


    if q_tc_interp.five_STRAwell(j,1)>0

        tc.abate.five_STRAwell(j)=0;
        mac.abate.five_STRAwell(j)=0;

    else
       tc.abate.five_STRAwell(j)=pchip(q_tc_interp.five_STRAwell(j,:),...
          [tc_interp.five_STRAwell(j,1:find(q_tc_interp.five_STRAwell(j,:)<
          max(q_tc_interp.five_STRAwell(j,:)),1,'last'))...
        tc_interp.five_STRAwell(j,find(q_tc_interp.five_STRAwell(j,:)<
        max(q_tc_interp.five_STRAwell(j,:)),1,'last'))
        +.00001],initial_abate_inches.five_STRAwell(j));

       mac.abate.five_STRAwell(j)=pchip(q_i.five_STRAwell(j,:),p_i.five_STRAwell(j,:),
       initial_abate_inches.five_STRAwell(j));

    end

tc.trade.five_STRAwell(j)=pchip(q_tc_interp.five_STRAwell(j,:),...
    [tc_interp.five_STRAwell(j,1:find(q_tc_interp.five_STRAwell(j,:)<
    max(q_tc_interp.five_STRAwell(j,:)),1,'last'))...
    tc_interp.five_STRAwell(j,find(q_tc_interp.five_STRAwell(j,:)<
    max(q_tc_interp.five_STRAwell(j,:)),1,'last'))+.00001],
    reduced_allocation_inches.five_STRAwell(j));

```

```
 else
     tc.abate.five_STRAwell(j)=0;
     tc.trade.five_STRAwell(j)=0;

 end


market_cost_trade.five_STRAwell(j)= inches_traded.five_STRAwell(j)*mac.tax.five_STRAwell(j);

welfare_gain.five_STRAwell(j) = (tc.abate.five_STRAwell(j)- tc.trade.five_STRAwell(j)
+ market_cost_trade.five_STRAwell(j))*five_STRAwell_acres(j,39);

if welfare_gain.five_STRAwell(j)<0
    welfare_gain.five_STRAwell(j)=0;
end

end

% Scale the sample estimates back to population estimates by well numbers
sum_gain.five_STRAwell(i)=sum(welfare_gain.five_STRAwell)*num_wells/size(five_STRAwell_wells,1);

% Scale the sample estimates back to population estimates by area
sum_gain_area.five_STRAwell(i)=sum(welfare_gain.five_STRAwell)
*sum(wells{2})/tot_area.five_STRAwell(i);

sum_gain_area.five_STRAwell_adj(i)=sum(welfare_gain.five_STRAwell(1:size(sample_s,1)))
*sum(wells{2}(small))/tot_area.well_s(i)...
+sum(welfare_gain.five_STRAwell(1+size(sample_s,1):size(sample_s,1)+size(sample_m,1)))
*sum(wells{2}(medium))/tot_area.well_m(i)...
+sum(welfare_gain.five_STRAwell(size(sample_s,1)+size(sample_m,1):size(sample_f,1)))
*sum(wells{2}(large))/tot_area.well_l(i);

% Get other characteristics of the samples
tot_area.five_STRAwell(i)=sum(wells{2}(sample_f));
tot_lift.well_five_STRAwell(i)=mean(lift(sample_f));
tot_yield.well_five_STRAwell(i)=mean(yield(sample_f));

% The number of buyers, sellers and outsiders.
wbuy.erfive_STRAwell(i)=sum(market_cost_trade.five_STRAwell<0);
wout.erfive_STRAwell(i)=sum(market_cost_trade.five_STRAwell==0);
wsell.erfive_STRAwell(i)=sum(market_cost_trade.five_STRAwell>0);

```

```
% The percentage of buyers, sellers and outsiders.
wbuy.erfive_STRAwell_r(i)=wbuy.erfive_STRAwell(i)/size(sample_f,1);
wout.erfive_STRAwell_r(i)=wout.erfive_STRAwell(i)/size(sample_f,1);
wsell.erfive_STRAwell_r(i)=wsell.erfive_STRAwell(i)/size(sample_f,1);

% People who sell all their permits and move to dryland
wsell.all_five_STRAwell(i)=sum(inches_traded.five_STRAwell'
≥current_allocation(sample_f))/size(sample_f,1);

% Amount of permits traded
wsell.traded_five_STRAwell(i)=sum(abs(inches_traded.five_STRAwell)
*five_STRAwell_acres(:,39))/2;

end
toc

% save it
```

Farm-base Random Sampling

In a draw of 5 percent farms, select the sampled farms by random number first.Then, include all the wells belonged to these farms. Use this code to generate the optimal permit prices, irrigation water allocation, and calculate the cost saving in abatement. Record these information for each single draw and save these data for further analysis. This is an example for 5 percent farm-base random sampling. I also replaced the ratio in the 10^{th} line in this code with 0.01 to 0.04 to get 1 percent to 4 percent samples.

```
1  % There are 4525 farms in total.
2  num_farms=4525;
3
4  % Get the id for farms which is the first one registered by a single owner ID.
5  [ans1, ans2, ans3]=unique(wells{22},'first');
6  firstwell=sort(ans2);
7
8  tic
9  % To get 5 percent samples
10 ratio=0.05;
11
12 % Take 1000 draws.
13 draws=1000;
14 num_sample=round(num_farms*ratio);
15
16 %set seed at 1
17 rand('seed',1);
18
19 for i=1:draws
20 clear sample_f
21
22 r=zeros(num_farms,1);
23 raw_sample=zeros(num_sample,1);
24 sample=zeros(num_wells,1);
25 selected=zeros(num_wells,1);
26 r=sortrows([round(num_farms.*rand(num_farms,1)) unique(wells{22})]);
27
28 % Get the farms to be sampled in this draw.
```

```
raw_sample=sort(r(1:num_sample,2));
% Get the wells belonging to these farms
for j=1:num_sample
  sample=sample+(wells{22}==raw_sample(j));
selected=sortrows(linspace(1,num_wells,num_wells)'.*sample,-1);
sample_f=sort(selected(1:find(selected>0,1,'last')));
end
% sample_f is the ID for the wells included into this draw.
ran5_value=value(sample_f,:);
ran5_inches=inches(sample_f,:);
ran5_acres_irrg=acres(sample_f,:);
ran5_acres=acres(sample_f,:) + dryland(sample_f,:);
ran5_well_id=well_id(sample_f,:);
ran5_dist_stream=dist_stream(sample_f,:);
ran5_dist_wells=dist_well(sample_f,:);
ran5_yield=yield(sample_f,:);
ran5_lift=lift(sample_f,:);
ran5_soil=soil(sample_f,:);
ran5_cert_acres= cert_acres(sample_f);
ran5_wells= wells{1}(sample_f);
ran5_pump_yield=pump_yield(sample_f);
ran5_soil_types=soil_types(sample_f);
for j=1:size(ran5_wells,1)
  p2.ran5(j,:)=p2.basin(sample_f(j),:);
  q2.ran5(j,:)=q2.basin(sample_f(j),:);
initial_abate.ran5(j)= max(ran5_inches(j,39)-current_allocation(sample_f(j)),0)*ran5_acres(j,39);
initial_abate_inches.ran5(j)= max(ran5_inches(j,39)-current_allocation(sample_f(j)),0);
temp_wells.ran5 = size(ran5_wells,1);
current_use.ran5(j)=min(current_allocation(sample_f(j)),ran5_inches(j,39));
free_ride.ran5(j)=current_allocation(sample_f(j))-current_use.ran5(j);
free_ride.ran5(find(ran5_inches(j,39)==0))=0;
ran5_acres_i(j)=ran5_acres(j,39);
```

```
end
for j=1:size(ran5_wells,1)
  if   p2.ran5(j,39)>0
        loc.start=find(p2.ran5(j,:)>0.0001,1,'first')-1;%last zero value
       %first zero gradient
        loc.end=(find(p2.ran5(j,:)>0.0001 & gradient(p2.ran5(j,:))<0.1,1,'first')-1);
        q0=q2.ran5(j,loc.start);%find baseline for normalization
        q_temp=q2.ran5(j,loc.start:loc.end);%pull out relevant points
        %normalize to origin
        q_temp=[q_temp-ones(size(q_temp)).*q0 ran5_inches(j,39) ran5_inches(j,39)+.00001];
       %adjust for slack constraints and initial water use
        q_temp=q_temp+ones(size(q_temp)).*free_ride.ran5(j);
        p_temp=[p2.ran5(j,loc.start:loc.end) p2.ran5(j,loc.end) 100];
        %the problem is below
q_junk.ran5(j,:)=linspace(min(q_temp),max(q_temp)- 0.01001,48);
p_i.ran5(j,:)=[pchip(q_temp,p_temp,q_junk.ran5(j,:)) p2.ran5(j,loc.end) 100];
q_i.ran5(j,:)=[q_junk.ran5(j,:) max(ran5_inches(j,39),current_allocation(sample_f(j)))
max(ran5_inches(j,39),current_allocation(sample_f(j)))+0.0001];
p_i.ran5(j,:) = p_i.ran5(j,:) + linspace(0,1e-6,size(p_i.ran5(j,:),2));
  elseif p2.ran5(j,39)<=0
       p_i.ran5(j,:)= zeros(1,50);
       p_i.ran5(j,1)= 1e10;
       q_i.ran5(j,:)= zeros(1,50);
  end
end
```

```


%find the mac (tax) where inches abated equals abatement under NRD allocations
tax.ran5(i) = fzero(@(u) sum(pigou(u,temp_wells.ran5,p_i.ran5,q_i.ran5,ran5_acres_i))
-sum(initial_abate.ran5),10);

mac.tax.ran5 = tax.ran5(i)*ones(size(ran5_wells,1),1); %vector of macs

for j=1:size(ran5_wells,1)


    if p_i.ran5(j,50)>0
%find acre inches reduced under equalized mac
 reduced_allocation.ran5(j) = pchip(p_i.ran5(j,:),q_i.ran5(j,:),
 mac.tax.ran5(j))*ran5_acres(j,39);
 %find inches per acre reduced under equalized mac
 reduced_allocation_inches.ran5(j)=pchip(p_i.ran5(j,:),q_i.ran5(j,:),mac.tax.ran5(j));

    elseif p_i.ran5(j,50)<=0

 reduced_allocation.ran5(j) =0; %for dryland wells

 reduced_allocation_inches.ran5(j)= 0; %for dryland wells

    end

acreinches_traded.ran5(j)=reduced_allocation.ran5(j)-initial_abate.ran5(j);

temp_free.ran5(j)=free_ride.ran5(j)*ran5_acres(j,39);

inches_traded.ran5(j)=reduced_allocation_inches.ran5(j)-initial_abate_inches.ran5(j);

end


for j=1:size(ran5_wells,1)


 if q_i.ran5(j,50)>0
q_tc_interp.ran5(j,:)=linspace(min(q_i.ran5(j,:)),max(q_i.ran5(j,:)),101);

tc_interp.ran5(j,:) = pchip(q_i.ran5(j,:),p_i.ran5(j,:),q_tc_interp.ran5(j,:))
```

```
*triu(ones(101))/100*(max(q_i.ran5(j,:))-min(q_i.ran5(j,:)));


    if q_tc_interp.ran5(j,1)>0

        tc.abate.ran5(j)=0;
        mac.abate.ran5(j)=0;

    else
       tc.abate.ran5(j)=pchip(q_tc_interp.ran5(j,:),...
           [tc_interp.ran5(j,1:find(q_tc_interp.ran5(j,:)<max(q_tc_interp.ran5(j,:)),1,'last'))...
           tc_interp.ran5(j,find(q_tc_interp.ran5(j,:)<max(q_tc_interp.ran5(j,:)),1,'last'))
           +.00001],initial_abate_inches.ran5(j));

       mac.abate.ran5(j)=pchip(q_i.ran5(j,:),p_i.ran5(j,:),initial_abate_inches.ran5(j));

    end


tc.trade.ran5(j)=pchip(q_tc_interp.ran5(j,:),...
    [tc_interp.ran5(j,1:find(q_tc_interp.ran5(j,:)<max(q_tc_interp.ran5(j,:)),1,'last'))...
    tc_interp.ran5(j,find(q_tc_interp.ran5(j,:)<max(q_tc_interp.ran5(j,:)),1,'last'))+.00001],
    reduced_allocation_inches.ran5(j));

 else
     tc.abate.ran5(j)=0;
     tc.trade.ran5(j)=0;

 end


market_cost_trade.ran5(j)= inches_traded.ran5(j)*mac.tax.ran5(j);

welfare_gain.ran5(j) = (tc.abate.ran5(j)- tc.trade.ran5(j) + market_cost_trade.ran5(j))
*ran5_acres(j,39);

if welfare_gain.ran5(j)<0

    welfare_gain.ran5(j)=0;
end


```

```
end

% Scale the sample estimates back to population estimates
sum_gain.ran5(i)=sum(welfare_gain.ran5(1:size(ran5_wells,1)))*num_wells/size(ran5_wells,1);
uni_wellid=unique((wells{58}(sample_f).*sample_f));
tot_area.ran5(i)=sum(wells{55}(uni_wellid(2:sum(wells{58}(sample_f)))));
sum_gain_area.ran5(i)=sum(welfare_gain.ran5(1:size(ran5_wells,1)))
*sum(wells{55}(firstwell))/tot_area.ran5(i);

% Get other characteristics of the samples
tot_area.ran5(i)=sum(wells{2}(sample_f));
tot_lift.ran5(i)=mean(ran5_lift(1:size(sample_f,1)));
tot_yield.ran5(i)=mean(ran5_yield(1:size(sample_f,1)));

% The number of buyers, sellers and outsiders.
buy.er_ran5(i)=sum(market_cost_trade.ran5(1:size(sample_f,1))<0);
out.er_ran5(i)=sum(market_cost_trade.ran5(1:size(sample_f,1))==0);
sell.er_ran5(i)=sum(market_cost_trade.ran5(1:size(sample_f,1))>0);

% The percentage of buyers, sellers and outsiders.
buy.er_ran5_r(i)=buy.er_ran5(i)/size(sample_f,1);
out.er_ran5_r(i)=out.er_ran5(i)/size(sample_f,1);
sell.er_ran5_r(i)=sell.er_ran5(i)/size(sample_f,1);

% People who sell all their permits and move to dryland
sell.all_ran5(i)=sum(inches_traded.ran5(1:size(sample_f,1))'
≥current_allocation(sample_f))/size(sample_f,1);

% Amount of permits traded
sell.traded_ran5(i)=sum(abs(inches_traded.ran5(1:size(sample_f,1)))
*ran5_acres(:,39))/2;

end

toc

% save it
```

Farm-base Systematic Sampling

I used the following code to systematically sample farms by area.

```
% Systematically sampling based on total acreage per farm

% Get the id for farms which is the first one registered by a single owner ID.
[ans1, ans2, ans3]=unique(wells{22},'first');
firstwell=sort(ans2);

% To get 5 percent samples
ratio=0.05;

% Take 1000 draws.
draws=1000;

tic
num_sample=round(num.farms*ratio);

%set seed at 1
rand('seed',1);

for i=1:draws
clear sample_f
sample=zeros(num_wells,1);
selected=zeros(num_wells,1);
r2=rand(1+num.farms,1);

% Starting point
r_start=ceil(r2(1)*(num.farms-1/ratio*(num_sample-1)));
sample_r=linspace(r_start,r_start+1/ratio*(num_sample-1),num_sample)';
r=sortrows([wells{55}(firstwell) unique(wells{22})]);

% Farms included into this draw
raw_sample=sort(r(sample_r,2));

% Get all the wells belonging to these farms
for j=1:num_sample
    sample=sample+(wells{22}==raw_sample(j));
selected=sortrows(linspace(1,num_wells,num_wells)'.*sample,-1);
```

```
sample_f=sort(selected(1:find(selected>0,1,'last')));
end

farmsys_area_value=value(sample_f,:);
farmsys_area_inches=inches(sample_f,:);
farmsys_area_acres_irrg=acres(sample_f,:);
farmsys_area_acres=acres(sample_f,:) + dryland(sample_f,:);
farmsys_area_well_id=well_id(sample_f,:);
farmsys_area_dist_stream=dist_stream(sample_f,:);
farmsys_area_dist_wells=dist_well(sample_f,:);
farmsys_area_yield=yield(sample_f,:);
farmsys_area_lift=lift(sample_f,:);
farmsys_area_soil=soil(sample_f,:);
farmsys_area_cert_acres= cert_acres(sample_f);
farmsys_area_wells= wells{1}(sample_f);
farmsys_area_pump_yield=pump_yield(sample_f);
farmsys_area_soil_types=soil_types(sample_f);


for j=1:size(farmsys_area_wells,1)

  p2.farmsys_area(j,:)=p2.basin(sample_f(j),:);
  q2.farmsys_area(j,:)=q2.basin(sample_f(j),:);

initial_abate.farmsys_area(j)= max(farmsys_area_inches(j,39)
-current_allocation(sample_f(j)),0)*farmsys_area_acres(j,39);

initial_abate_inches.farmsys_area(j)= max(farmsys_area_inches(j,39)
-current_allocation(sample_f(j)),0);


temp_wells.farmsys_area = size(farmsys_area_wells,1);

current_use.farmsys_area(j)=min(current_allocation(sample_f(j)),
farmsys_area_inches(j,39));
free_ride.farmsys_area(j)=current_allocation(sample_f(j))-current_use.farmsys_area(j);
free_ride.farmsys_area(find(farmsys_area_inches(j,39)==0))=0;
farmsys_area_acres_i(j)=farmsys_area_acres(j,39);

end
```

```
for j=1:size(farmsys_area_wells,1)
  if   p2.farmsys_area(j,40)>0
       loc.start=find(p2.farmsys_area(j,:)>0.0001,1,'first')−1;%last zero value
       loc.end=(find(p2.farmsys_area(j,:)>0.0001 & gradient(p2.farmsys_area(j,:))<0.1,
       1,'first')−1);%first zero gradient
       q0=q2.farmsys_area(j,loc.start);%find baseline for normalization
       q_temp=q2.farmsys_area(j,loc.start:loc.end);%pull out relevant points
       q_temp=[q_temp−ones(size(q_temp)).*q0 farmsys_area_inches(j,39)
       farmsys_area_inches(j,39)+.00001];%normalize to origin
      %adjust for slack constraints and initial water use
       q_temp=q_temp+ones(size(q_temp)).*free_ride.farmsys_area(j);
       p_temp=[p2.farmsys_area(j,loc.start:loc.end) p2.farmsys_area(j,loc.end) 100];
       %the problem is below
q_junk.farmsys_area(j,:)=linspace(min(q_temp),max(q_temp)− 0.01001,48);
p_i.farmsys_area(j,:)=[pchip(q_temp,p_temp,q_junk.farmsys_area(j,:))
p2.farmsys_area(j,loc.end) 100];
q_i.farmsys_area(j,:)=[q_junk.farmsys_area(j,:) max(farmsys_area_inches(j,39),
current_allocation(sample_f(j))) max(farmsys_area_inches(j,39),
current_allocation(sample_f(j)))+0.0001];
p_i.farmsys_area(j,:) = p_i.farmsys_area(j,:) + linspace(0,1e−6,size(p_i.farmsys_area(j,:),2));
  elseif p2.farmsys_area(j,40)≤0
      p_i.farmsys_area(j,:)= zeros(1,50);
      p_i.farmsys_area(j,1)= 1e10;
      q_i.farmsys_area(j,:)= zeros(1,50);
  end
```

```
end

%find the mac (tax) where inches abated equals abatement under NRD allocations
tax.farmsys_area(i) = fzero(@(u) sum(pigou(u,temp_wells.farmsys_area,
p_i.farmsys_area,q_i.farmsys_area,
farmsys_area_acres_i))-sum(initial_abate.farmsys_area),10);

mac.tax.farmsys_area = tax.farmsys_area(i)*ones(size(farmsys_area_wells,1),1); %vector of macs

for j=1:size(farmsys_area_wells,1)

    if p_i.farmsys_area(j,50)>0

 reduced_allocation.farmsys_area(j) = pchip(p_i.farmsys_area(j,:),q_i.farmsys_area(j,:),
 mac.tax.farmsys_area(j))*farmsys_area_acres(j,39); %find acre inches reduced under equalized mac
 reduced_allocation_inches.farmsys_area(j)=pchip(p_i.farmsys_area(j,:),q_i.farmsys_area(j,:),
 mac.tax.farmsys_area(j)); %find inches per acre reduced under equalized mac

    elseif p_i.farmsys_area(j,50)≤0

 reduced_allocation.farmsys_area(j) =0; %for dryland wells

 reduced_allocation_inches.farmsys_area(j)= 0; %for dryland wells

    end

acreinches_traded.farmsys_area(j)=reduced_allocation.farmsys_area(j)-initial_abate.farmsys_area(j);

temp_free.farmsys_area(j)=free_ride.farmsys_area(j)*farmsys_area_acres(j,39);

inches_traded.farmsys_area(j)=reduced_allocation_inches.farmsys_area(j)
-initial_abate_inches.farmsys_area(j);

end


for j=1:size(farmsys_area_wells,1)
```

```
 if q_i.farmsys_area(j,50)>0
q_tc_interp.farmsys_area(j,:)=linspace(min(q_i.farmsys_area(j,:)),max(q_i.farmsys_area(j,:)),101);

tc_interp.farmsys_area(j,:) = pchip(q_i.farmsys_area(j,:),p_i.farmsys_area(j,:),
q_tc_interp.farmsys_area(j,:))
*triu(ones(101))/100*(max(q_i.farmsys_area(j,:))-min(q_i.farmsys_area(j,:)));


    if q_tc_interp.farmsys_area(j,1)>0

        tc.abate.farmsys_area(j)=0;
        mac.abate.farmsys_area(j)=0;

    else
       tc.abate.farmsys_area(j)=pchip(q_tc_interp.farmsys_area(j,:),...
           [tc_interp.farmsys_area(j,1:find(q_tc_interp.farmsys_area(j,:)<
           max(q_tc_interp.farmsys_area(j,:)),1,'last'))...
           tc_interp.farmsys_area(j,find(q_tc_interp.farmsys_area(j,:)<
           max(q_tc_interp.farmsys_area(j,:)),1,'last'))+.00001],
           initial_abate_inches.farmsys_area(j));

       mac.abate.farmsys_area(j)=pchip(q_i.farmsys_area(j,:),p_i.farmsys_area(j,:),
       initial_abate_inches.farmsys_area(j));

    end


tc.trade.farmsys_area(j)=pchip(q_tc_interp.farmsys_area(j,:),...
    [tc_interp.farmsys_area(j,1:find(q_tc_interp.farmsys_area(j,:)<
    max(q_tc_interp.farmsys_area(j,:)),1,'last'))...
    tc_interp.farmsys_area(j,find(q_tc_interp.farmsys_area(j,:)<
    max(q_tc_interp.farmsys_area(j,:)),1,'last'))+.00001],
    reduced_allocation_inches.farmsys_area(j));

 else
     tc.abate.farmsys_area(j)=0;
     tc.trade.farmsys_area(j)=0;

 end
```

```
market_cost_trade.farmsys_area(j)= inches_traded.farmsys_area(j)*mac.tax.farmsys_area(j);

welfare_gain.farmsys_area(j) = (tc.abate.farmsys_area(j)- tc.trade.farmsys_area(j)
+ market_cost_trade.farmsys_area(j))*farmsys_area_acres(j,39);

if welfare_gain.farmsys_area(j)<0

    welfare_gain.farmsys_area(j)=0;
end


end

% Scale the sample estimates back to population estimates by well numbers.
sum_gain.farmsys_area(i)=sum(welfare_gain.farmsys_area(1:size(farmsys_area_wells,1)))
*num_wells/size(farmsys_area_wells,1);

% Get other characteristics of the samples
tot_area.farmsys_area(i)=sum(wells{2}(sample_f));
tot_lift.farmsys_area(i)=mean(farmsys_area_lift(1:size(sample_f,1)));
tot_yield.farmsys_area(i)=mean(farmsys_area_yield(1:size(sample_f,1)));

% Scale the sample estimates back to population estimates by acres.
sum_gain_area.farmsys_area(i)=sum(welfare_gain.farmsys_area(1:size(farmsys_area_wells,1)))
*sum(wells{55}(firstwell))/tot_area.farmsys_area(i);

% The number of buyers, sellers and outsiders.
    buy.er_farmsys_area(i)=sum(market_cost_trade.farmsys_area(1:size(sample_f,1))<0);
out.er_farmsys_area(i)=sum(market_cost_trade.farmsys_area(1:size(sample_f,1))==0);
sell.er_farmsys_area(i)=sum(market_cost_trade.farmsys_area(1:size(sample_f,1))>0);

% The percentage of buyers, sellers and outsiders.
    buy.er_farmsys_area_r(i)=buy.er_farmsys_area(i)/size(sample_f,1);
out.er_farmsys_area_r(i)=out.er_farmsys_area(i)/size(sample_f,1);
sell.er_farmsys_area_r(i)=sell.er_farmsys_area(i)/size(sample_f,1);

% People who sell all their permits and move to dryland
sell.all_farmsys_area(i)=sum(inches_traded.farmsys_area(1:size(sample_f,1))'
≥current_allocation(sample_f))/size(sample_f,1);

% Amount of permits traded
sell.traded_farmsys_area(i)=sum(abs(inches_traded.farmsys_area(1:size(sample_f,1)))
```

```
*farmsys_area_acres(:,39))/2;

end

toc

% save it
```

If I replace line 16 to line 38 by the following code, I can systematically select farms by their owner ID. However, the result is very close to systematic sampling by area.

```
%set seed at 1
rand('seed',1);

for i=1:draws
    sample=zeros(num_wells,1);
selected=zeros(num_wells,1);
r2=rand(1+num.farms,1);
r_start=ceil(r2(1)*(num.farms-1/ratio*(num_sample-1)));

sample_r=linspace(r_start,r_start+1/ratio*(num_sample-1),num_sample)';

r=sortrows([round(num.farms.*r2(2:1+num.farms))
linspace(1,num.farms,num.farms)' unique(wells{22})]);
raw_sample=sort(r(sample_r,3));

for j=1:num_sample
    sample=sample+(wells{22}==raw_sample(j));
selected=sortrows(linspace(1,num_wells,num_wells)'.*sample,-1);
sample_f=sort(selected(1:find(selected>0,1,'last')));
end
```

Farm-base Stratified Sampling

The strata used for this code are in Table A-2.

```
% Get the id for farms which is the first one registered by a single owner ID.
[ans1, ans2, ans3]=unique(wells{22},'first');
firstwell=sort(ans2);

% Define small, medium and large wells.
small=(wells{55}≤160);
medium=(wells{55}>160 & wells{55}≤320);
large=(wells{55}>320);

% Define small, medium and large farm.
smallfarm=unique(wells{22}(wells{55}≤160));
mediumfarm=unique(wells{22}(wells{55}>160 & wells{55}≤320));
largefarm=unique(wells{22}(wells{55}>320));

% The number of farms in each stratum
num.smallfarm=size(smallfarm,1);
num.mediumfarm=size(mediumfarm,1);
num.largefarm=size(largefarm,1);
gbg=wells{55}.*wells{58};
tic

% To get 5 percent samples
ratio=0.05;

% Take 1000 draws.
draws=1000;

num_sample=round(num.farms*ratio);
num_sample_smallfarm=round(num.smallfarm*ratio);
num_sample_mediumfarm=round(num.mediumfarm*ratio);
num_sample_largefarm=round(num.largefarm*ratio);

%set seed at 1
rand('seed',1);

for i=1:draws
```

```
clear sample_f
sample_s=zeros(num_wells,1);
selected_s=zeros(num_wells,1);
sample_m=zeros(num_wells,1);
selected_m=zeros(num_wells,1);
sample_l=zeros(num_wells,1);
selected_l=zeros(num_wells,1);

% Get random farms from each stratum.
r3=rand(num.smallfarm+num.mediumfarm+num.largefarm,1);
r_s=sortrows([round(num.smallfarm.*r3(1:num.smallfarm)) smallfarm]);
raw_sample_s(:,i)=sort(r_s(1:num_sample_smallfarm,2));
r_m=sortrows([round(num.mediumfarm.*r3(num.smallfarm+1: num.smallfarm
+num.mediumfarm)) mediumfarm]);
raw_sample_m(:,i)=sort(r_m(1:num_sample_mediumfarm,2));
r_l=sortrows([round(num.largefarm.*r3(num.smallfarm+num.mediumfarm+1: num.smallfarm
+num.mediumfarm+num.largefarm)) largefarm]);
raw_sample_l(:,i)=sort(r_l(1:num_sample_largefarm,2));

% Wells belonged to these farms
for j=1:num_sample_smallfarm
  included_s=(wells{22}==raw_sample_s(j,i));
  sample_s=sample_s+included_s;
  selected_s=sortrows(linspace(1,num_wells,num_wells)'.*sample_s,-1);
sample_fs=sort(selected_s(1:find(selected_s>0,1,'last')));
end
for j=1:num_sample_mediumfarm
  included_m=(wells{22}==raw_sample_m(j,i));
  sample_m=sample_m+included_m;
  selected_m=sortrows(linspace(1,num_wells,num_wells)'.*sample_m,-1);
sample_fm=sort(selected_m(1:find(selected_m>0,1,'last')));
end
for j=1:num_sample_largefarm
  included_l=(wells{22}==raw_sample_l(j,i));
  sample_l=sample_l+included_l;
  selected_l=sortrows(linspace(1,num_wells,num_wells)'.*sample_l,-1);
sample_fl=sort(selected_l(1:find(selected_l>0,1,'last')));
end

% Combine wells from each stratum into a sample for ith draw.
sample_f=[sample_fs; sample_fm; sample_fl];
```

```
five_STRAfarm_value=value(sample_f,:);
five_STRAfarm_inches=inches(sample_f,:);
five_STRAfarm_acres_irrg=acres(sample_f,:);
five_STRAfarm_acres=acres(sample_f,:); + dryland(sample_f,:);
five_STRAfarm_well_id=well_id(sample_f,:);
five_STRAfarm_dist_stream=dist_stream(sample_f,:);
five_STRAfarm_dist_wells=dist_well(sample_f,:);
five_STRAfarm_yield=yield(sample_f,:);
five_STRAfarm_lift=lift(sample_f,:);
%five_STRAfarm_soil=STRAfarm_soil(sample(:,i),:);
five_STRAfarm_cert_acres= cert_acres(sample_f);
five_STRAfarm_wells= wells{1}(sample_f);

for j=1:size(sample_f,1)
p2.five_STRAfarm(j,:)=wrev(a.p.basin(sample_f(j),:));

q2.five_STRAfarm(j,:)=wrev(a.q.basin(sample_f(j),:));

initial_abate.five_STRAfarm(j)= max(five_STRAfarm_inches(j,39)
-current_allocation(sample_f(j)),0)*five_STRAfarm_acres(j,39);

initial_abate_inches.five_STRAfarm(j)= max(five_STRAfarm_inches(j,39)
-current_allocation(sample_f(j)),0);

temp_wells.five_STRAfarm = size(five_STRAfarm_wells,1);

current_use.five_STRAfarm(j)=min(current_allocation(sample_f(j)),
five_STRAfarm_inches(j,39));
free_ride.five_STRAfarm(j)=current_allocation(sample_f(j))
-current_use.five_STRAfarm(j);
free_ride.five_STRAfarm(find(five_STRAfarm_inches(j,39)==0))=0;
five_STRAfarm_acres_i(j)=five_STRAfarm_acres(j,39);

end

for j=1:size(sample_f,1)

  if   p2.five_STRAfarm(j,40)>0
      %last zero value
       loc.start=find(p2.five_STRAfarm(j,:)>0.0001,1,'first')-1;
      %first zero gradient
       loc.end=(find(p2.five_STRAfarm(j,:)>0.0001 & gradient(p2.five_STRAfarm(j,:))<0.1,
```

```
        1,'first')-1);
        %find baseline for normalization
        q0=q2.five_STRAfarm(j,loc.start);

        q_temp=q2.five_STRAfarm(j,loc.start:loc.end);%pull out relevant points

        q_temp=[q_temp-ones(size(q_temp)).*q0 five_STRAfarm_inches(j,39)
        five_STRAfarm_inches(j,39)+.00001];%normalize to origin

       %adjust for slack constraints and initial water use
        q_temp=q_temp+ones(size(q_temp)).*free_ride.five_STRAfarm(j);

        p_temp=[p2.five_STRAfarm(j,loc.start:loc.end) p2.five_STRAfarm(j,loc.end) 100];

        %the problem is below
q_junk.five_STRAfarm(j,:)=linspace(min(q_temp),max(q_temp)- 0.01001,48);

p_i.five_STRAfarm(j,:)=[pchip(q_temp,p_temp,q_junk.five_STRAfarm(j,:))
p2.five_STRAfarm(j,loc.end) 100];

q_i.five_STRAfarm(j,:)=[q_junk.five_STRAfarm(j,:) max(five_STRAfarm_inches(j,39),
current_allocation(sample_f(j))) max(five_STRAfarm_inches(j,39),
current_allocation(sample_f(j)))+0.0001];

p_i.five_STRAfarm(j,:)=p_i.five_STRAfarm(j,:)+linspace(0,1e-6,size(p_i.five_STRAfarm(j,:),2));

  elseif p2.five_STRAfarm(j,40)<=0
       p_i.five_STRAfarm(j,:)= zeros(1,50);
       p_i.five_STRAfarm(j,1)= 1e10;
       q_i.five_STRAfarm(j,:)= zeros(1,50);
   end

end


%find the mac (tax) where inches abated equals abatement under NRD allocations
tax.five_STRAfarm(i) = fzero(@(u) sum(pigou(u,temp_wells.five_STRAfarm,
p_i.five_STRAfarm,q_i.five_STRAfarm,
five_STRAfarm_acres_i))-sum(initial_abate.five_STRAfarm),10);

%vector of macs
mac.tax.five_STRAfarm = tax.five_STRAfarm(i)*ones(size(five_STRAfarm_wells,1),1);
```

```
for j=1:size(sample_f,1)

    if p_i.five_STRAfarm(j,50)>0
%find acre inches reduced under equalized mac
 reduced_allocation.five_STRAfarm(j) = pchip(p_i.five_STRAfarm(j,:),q_i.five_STRAfarm(j,:),
 mac.tax.five_STRAfarm(j))*five_STRAfarm_acres(j,39);
 reduced_allocation_inches.five_STRAfarm(j)=pchip(p_i.five_STRAfarm(j,:),q_i.five_STRAfarm(j,:),
 mac.tax.five_STRAfarm(j)); %find inches per acre reduced under equalized mac

    elseif p_i.five_STRAfarm(j,50)≤0

 reduced_allocation.five_STRAfarm(j) =0; %for dryland wells

 reduced_allocation_inches.five_STRAfarm(j)= 0; %for dryland wells

    end

acreinches_traded.five_STRAfarm(j)=reduced_allocation.five_STRAfarm(j)
-initial_abate.five_STRAfarm(j);

temp_free.five_STRAfarm(j)=free_ride.five_STRAfarm(j)*five_STRAfarm_acres(j,39);

inches_traded.five_STRAfarm(j)=reduced_allocation_inches.five_STRAfarm(j)
-initial_abate_inches.five_STRAfarm(j);

end

for j=1:size(sample_f,1)

 if q_i.five_STRAfarm(j,50)>0
q_tc_interp.five_STRAfarm(j,:)=linspace(min(q_i.five_STRAfarm(j,:)),
max(q_i.five_STRAfarm(j,:)),101);

tc_interp.five_STRAfarm(j,:) = pchip(q_i.five_STRAfarm(j,:),p_i.five_STRAfarm(j,:),
q_tc_interp.five_STRAfarm(j,:))
*triu(ones(101))/100*(max(q_i.five_STRAfarm(j,:))-min(q_i.five_STRAfarm(j,:)));

    if q_tc_interp.five_STRAfarm(j,1)>0

        tc.abate.five_STRAfarm(j)=0;
        mac.abate.five_STRAfarm(j)=0;
```

```
    else
       tc.abate.five_STRAfarm(j)=pchip(q_tc_interp.five_STRAfarm(j,:),...
           [tc_interp.five_STRAfarm(j,1:find(q_tc_interp.five_STRAfarm(j,:)<
           max(q_tc_interp.five_STRAfarm(j,:)),1,'last'))...
           tc_interp.five_STRAfarm(j,find(q_tc_interp.five_STRAfarm(j,:)<
           max(q_tc_interp.five_STRAfarm(j,:)),1,'last'))+.00001],
           initial_abate_inches.five_STRAfarm(j));

       mac.abate.five_STRAfarm(j)=pchip(q_i.five_STRAfarm(j,:),p_i.five_STRAfarm(j,:),
       initial_abate_inches.five_STRAfarm(j));

    end

tc.trade.five_STRAfarm(j)=pchip(q_tc_interp.five_STRAfarm(j,:),...
    [tc_interp.five_STRAfarm(j,1:find(q_tc_interp.five_STRAfarm(j,:)<
    max(q_tc_interp.five_STRAfarm(j,:)),1,'last'))...
    tc_interp.five_STRAfarm(j,find(q_tc_interp.five_STRAfarm(j,:)<
    max(q_tc_interp.five_STRAfarm(j,:)),1,'last'))
    +.00001],reduced_allocation_inches.five_STRAfarm(j));

 else
     tc.abate.five_STRAfarm(j)=0;
     tc.trade.five_STRAfarm(j)=0;

 end

market_cost_trade.five_STRAfarm(j)= inches_traded.five_STRAfarm(j)
*mac.tax.five_STRAfarm(j);

welfare_gain.five_STRAfarm(j) = (tc.abate.five_STRAfarm(j)- tc.trade.five_STRAfarm(j)
+ market_cost_trade.five_STRAfarm(j))*five_STRAfarm_acres(j,39);

if welfare_gain.five_STRAfarm(j)<0
       welfare_gain.five_STRAfarm(j)=0;
end

end

% Scale the sample estimates back to population estimates by well numbers
sum_gain.five_STRAfarm(i)=sum(welfare_gain.five_STRAfarm(1:size(sample_f,1)))
*num_wells/size(sample_f,1);
```

```
tot_area.five_STRAfarm(i)=sum(wells{2}(sample_f));

% Scale the sample estimates back to population estimates by acres
sum_gain_area.five_STRAfarm(i)=sum(welfare_gain.five_STRAfarm(1:size(sample_f,1)))
*sum(wells{55}(firstwell))/tot_area.five_STRAfarm(i);
sum_gain_area.five_STRAfarm_adj(i)=sum(welfare_gain.five_STRAfarm(1:size(sample_fs,1)))
*sum(gbg(small))/sum(wells{2}(sample_fs))...
    +sum(welfare_gain.five_STRAfarm(size(sample_fs,1)
    +1:(size(sample_fs,1)+size(sample_fm,1))))
    *sum(gbg(medium))/sum(wells{2}(sample_fm))...
    +sum(welfare_gain.five_STRAfarm((size(sample_fs,1)
    +size(sample_fm,1)+1):size(five_STRAfarm_wells,1)))
    *sum(gbg(large))/sum(wells{2}(sample_fl));

% Get other characteristics of the samples
tot_lift.five_STRAfarm(i)=mean(five_STRAfarm_lift(1:size(sample_f,1)));
tot_yield.five_STRAfarm(i)=mean(five_STRAfarm_yield(1:size(sample_f,1)));

% The number of buyers, sellers and outsiders.
    buy.er_five_STRAfarm(i)=sum(market_cost_trade.five_STRAfarm(1:size(sample_f,1))<0);
out.er_five_STRAfarm(i)=sum(market_cost_trade.five_STRAfarm(1:size(sample_f,1))==0);
sell.er_five_STRAfarm(i)=sum(market_cost_trade.five_STRAfarm(1:size(sample_f,1))>0);

% The percentage of buyers, sellers and outsiders.
    buy.er_five_STRAfarm_r(i)=buy.er_five_STRAfarm(i)/size(sample_f,1);
out.er_five_STRAfarm_r(i)=out.er_five_STRAfarm(i)/size(sample_f,1);
sell.er_five_STRAfarm_r(i)=sell.er_five_STRAfarm(i)/size(sample_f,1);

% People who sell all their permits and move to dryland
sell.all_five_STRAfarm(i)=sum(inches_traded.five_STRAfarm(1:size(sample_f,1))'
≥current_allocation(sample_f))/size(sample_f,1);

% Amount of permits traded
sell.traded_five_STRAfarm(i)=sum(abs(inches_traded.five_STRAfarm(1:size(sample_f,1)))
*five_STRAfarm_acres(:,39))/2;

end
toc

% save it
```

References

Ancev, Tihomir, Arthur L. Stoecker, Daniel E. Storm, and Michael J. White, "The Economics of Efficient Phosphorus Abatement in a Watershed," *Journal of Agricultural and Resource Economics*, 2006, *31*, 529–548.

Assael, Henry and John Keon, "Nonsampling vs. Sampling Errors in Survey Research," *Journal of Marketing*, 1982, *46*, 114–123.

Boggess, William, Ronald Lacewell, and David Zilberman, *Agricultural and Environmental Resource Economics*, New York, NY: Oxford University Press, 1993.

Braden, John B., Gary V. Johnson, Aziz Bouzaher, and David Miltz, "Optimal Spatial Management of Agricultural Pollution," *American Journal of Agricultural Economics*, 1989, *71*, 404–413.

Brozović, Nicholas, Janis M. Carey, and David L. Sunding, "Trading Activity in an Informal Agricultural Water Market: An Example from California," *Water Resources Update*, 2002, *121*, 3–16.

Chong, Howard and David Sunding, "Water Markets and Trading," *Annual Review of Environment and Resources*, 2006, *31*, 239–264.

Colby, B. G., K. Crandall, and D. B. Bush, "Water Right Transactions: Market Values and Price Dispersion," *Water Resources Research*, 1993, *29(6)*, 1565C1572.

Diao, Xinshen, Terry Roe, and Rachid Koukkali, "Economy-wide gains from decentralized water allocation in a spatially heterogenous agricultural economy," *Environment and Development Economics*, 2005, *10*, 249–269.

Diwakara, Halanaik and MG Chandrakanth, "Beating Negative Externality through Groundwater Recharge in India: A Resource Economic Analysis," *Environment and Development Economics*, 2007, *12*, 271C296.

FRIS, *Farm and Ranch Irrigation Survey*, Vol. 3, Washington, DC: Census of Agriculture U.S. Department of Commerce, Bureau of the Census, 1984.

———, *Farm and Ranch Irrigation Survey*, Vol. 3, Washington, DC: Census of Agriculture U.S. Department of Commerce, Bureau of the Census, 1988.

———, *Farm and Ranch Irrigation Survey*, Vol. 3, Washington, DC: Census of Agriculture U.S. Department of Commerce, Bureau of the Census, 2003.

Gonzalez-Alvarez, Yassert, Andrew G. Keller, and Jeffery D. Mullen, "Farm-level Irrigation and the Marginal Cost of Water Use; Evidence From Georgia," *Journal of Environmental Management*, 2006, *80*, 311–317.

Goodwin, Barry K., Ashok K. Mishra, and F. Ortal Magne, "What's Wrong With Our Models of Agricultural Land Values," *American Journal of Agricultural Economics*, 2003, *85(3)*, 744–752.

Hanley, Nick, Jason F. Shogren, and Ben White, *Environmental economics in theory and practice*, New York, NY: Palgrave MacMillan, 2002.

Hendricks, Nathan, *Estimating Irrigation Water Demand with a Multinomial Logit Selectivity Model*, Master's Thesis, Department of Agricultural Economics, Kansas State University, 2007.

Hinderlider, M.C, George S. Knapp, and Wardner G. Scott, "Public No. 696, Republican River Compact," 1942.

Ise, Sabrina and David L. Sunding, "Reallocating Water from Agriculture to the Environment under a Voluntary Purchase Program," *Review of Agricultural Economics*, 1998, *20*, 214–226.

Jenkins, Marion W., Jay R. Lund, Richard E. Howitt, Andrew J.Draper, Siwa M. Msangi, Stacy K. Tanaka, Randall S. Ritzema, and Guilherme F. Marques, "Optimization of Californias Water Supply System:Results and Insights," *Journal of Water Resources Planning and Management*, 2004, *130*, 271–280.

Katchova, A.L. and M.J. Miranda, "Two-Step Econometric Estimation of Farm Characteristics Affecting Marketing Contracts Decisions," *American Journal of Agricultural Economics*, 2004, *86(1)*, 88–102.

Khanna, Madhu, Richard Farnsworth, and Hayri Onal, "Targeting of CREP to Improve Water Quality: Determining Land Rental Offers with Endogenous Sediment Deposition Coefficients," *American Journal of Agricultural Economics*, 2003, *83*, 538–553.

Koundouri, Phoebe, Celine Nauges, and Vangelis Tzouvelekas, "Technology Adoption Under Production Uncertainty: Theory and Application to Irrigation Technology," *American Journal of Agricultural Economics*, 2006, *88*, 657–670.

Li, Qi and Jeffrey Scott Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, 2007.

Martin, Derrel, Ray Supalla, and Scott Nedved, *WaterOptimizer: Decision Support Tool for Deficit Irrigation Instruction Manual* Nebraska Cooperative Extension at University of Nebraska-Lincoln 2005.

McKusick, Vincent L., "State of Kansas v State of Nebraska and State of Colorado: Joint Motion of the States for the Entry of Proposed Consent Judgment and Approval and Adoption of Final Settlement Stipulation," 2002.

Moore, Michael R. and Ariel Dinar, "Water and Land as Quantity-Rationed Inputs in California Agriculture: Empirical Tests and Water Policy Implications," *Land Economics*, 1995, *71*, 445–461.

——, **Noel R. Gollehon, and Marc B. Carey**, "Multicrop Production Decisions in Western Irrigated Agriculture: The Role of Water Price," *American Journal of Agricultural Economics*, 1994, *76*, 859–874.

Morrison-Paul, Catherine, Richard Nehring, and David Banker, "Productivity, Economies, and Efficiency in U.S. Agriculture: A Look at Contracts," *American Journal of Agricultural Economics*, 2004, *86(5)*, 1308–1314.

Nebraska Department of Natural Resources and Lower Republican Natural Resource District, *Integrated Management Plan* 2008.

—— **and Middle Republican Natural Resource District**, *Integrated Management Plan* 2008.

—— **and Tri-Basin Natural Resource District**, *Integrated Management Plan* 2007.

—— **and Upper Republican Natural Resource District**, *Integrated Management Plan* 2008.

Nelson, Gerald C., "Introduction to The Special Issue on Spatial Analysis for Agricultural economists," *Agricultural Economics*, 2002, *27*, 197–200.

Palazzo, Amanda M., *Farm-level Impacts of Alternative Spatial Water Management Policies for the Protection of Instream Flows*, Master's Thesis, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 2009.

Pujol, Joan, Meri Raggi, and Davide Viaggi, "The Potential Impact of Markets for Irrigation Water in Italy and Spain: A Comparison of Two Study Areas," *The Australian Journal of Agricultural and Resources Economics*, 2006, *50*, 1467–8489.

Satti, Sudheer R. and Jennifer M Jacobs, "A GIS-based Model to Estimate the Regionally Distributed Drought Water Demand," *Agricultural Water Management*, 2004, *66*, 1–13.

Schaible, Glenn D., "Water Conservation Policy Analysis: An Interregional, Multi-Output, Primal-Dual Optimization Approach," *American Journal of Agricultural Economics*, 1997, *79*, 163–177.

Wockell, Edward L. and J. William Asher, *Educational Research*, New York: Macmillan: Prentice Hall, 1994.

Yang, Wanhong, Madhu Khanna, Richard Farnsworth, and Hayri Onal, "Integrating Economic, Environmental and GIS Modeling to Target Cost Effective Land Retirement in Multiple Watersheds," *Ecological Economics*, 2003, *46*, 249–267.