WESTERN REGIONAL RESEARCH PUBLICATION

W-2133

Benefits and Costs of Resource policies Affecting Public and Private Lands

Papers from the Annual Meeting

Waikoloa Village, Hawaii, February 17-20, 2008

Twenty-First Interim Report

June 2008

WESTERN REGIONAL RESEARCH PUBLICATION

W-2133
Benefits and Costs of Resource policies Affecting Public and Private Lands

Papers from the Annual Meeting
Waikoloa Village, Hawaii, February 17-20, 2008
Twenty-First Interim Report
June 2008

Compiled by
Kelly Giraud Cullen

Department of Resource Economics and Development
College of Life Science and Agriculture
University of New Hampshire
Durham, NH 03824

Table of Contents

NAME	AFFILIATION
William Breffle	Michigan Technical University
John Bergstrom	University of Georgia
Craig Bond	Colorado State University
Bonnie Colby	University of Arizona
Don Dennis	USDA Forest Service
Eric English	Stratus Consulting, Inc
Jerry Fletcher	West Virginia University
Kelly Giraud Cullen (president)	University of New Hampshire
Roger von Haefen (treasurer)	North Carolina State University
Robert Hearne	North Dakota State University
Joe Herriges	Iowa State University
John Hoehn	Michigan State University
Paul Jakus	Utah State University
Mike Kaplowitz	Michigan State University
Dan Lew	National Marine Fishery Service
John Loomis	Colorado State University
Frank Lupi (vice president)	Michigan State University
Vishakha Maskey (student)	West Virginia University
Doug Morris	University of New Hampshire
Dan McCollum	USDA, Forest Service
Klaus Moeltner	University of Nevada, Reno
Mary Riddel	University of Nevada, Las Vegas
Kim Rollins	University of Nevada, Reno
Randall Rosenberger	Oregon State University
Don Snyder (administrative advisor)	Utah State University
Brent Sohngen	Ohio State University

W2133 41st Annual Meeting

Hilton Waikoloa Village Resort on the Big Island of Hawaii, Feb 17-20, 2008

Sunday, February 17: Registration Opens

Monday, 8:00 AM–12:15 PM: W2133 Presidential Address and Welcome

Paper Session 1: Choosy Respondents: Choice Experiments and Choice Models

Chair: Michael KAPLOWITZ, Michigan State Univ

The Optimal Design: A Guide for Choice Experiment Practitioners. William BREFFLE, Michigan Technological University, Houghton, MI

A Discrete-Choice Model of Annual License Demand. Eric ENGLISH, Stratus Consulting Inc, Boulder, CO

Valuing Additional Protection of Steller Sea Lions Using a Choice Experiment-Based Stated Preference Method: Preliminary Results. Daniel K. LEW, Alaska Fisheries Science Center, National Marine Fisheries Service, Deattle, WA

A Choice Experiments Analysis of Stakeholder Preferences for Water Management Alternative in the Red River Basin. Bob HEARNE, North Dakota State University, Fargo, ND

Incorporating Random Coefficients and Alternative Specific Constraints into Discrete Choice Models: Implication for In-Sample Fir and Welfare Estimates. Roger H. VON HAEFEN, North Carolina State Univ., Raleigh

Meta-Analysis and Benefit Transfer: What do we Gain from Using Individual-level Data? Klaus MOELTNER, Univ Nevada Reno

Monday, 1:30 –5:45 PM: Paper Session 2A ~ W-2133/WRSA Joint Paper Session

Chair: Dan MCCOLLUM, U.S. Forest Service, Rocky Mtn Research Station, Ft. Collins, CO

A Panel-Mixed Logit Analysis of Colorado Corn Farmers' Stated Preferences for Private-Public Irrigation System Attributes. Craig BOND, Colorado State Univ., Fort Collins. Discussant: Roger VON HAEFEN, North Carolina State Univ., Raleigh

A Hybrid Individual-Zonal Travel Cost Model for Estimating the Consumer Surplus of Golfing in Colorado. John LOOMIS, Omer TADJION, Philip WATSON, Josh WILSON, Steve DAVIES, and Dawn THILMANY, Colorado State Univ., Fort Collins. Discussant: Robert HEARNE, North Dakota State Univ., Fargo

Selection Effects in Meta-Valuation Function Transfers. Randall ROSENBERGER, Oregon State Univ., Corvallis, and Robert J. JOHNSTON, Univ. Connecticut, Groton. Discussant: John BERGSTROM, Univ. Georgia, Athens

Economic Globalization and Resource Use: Do Growth, Investment, and Trade Encourage Water Use or Water Conservation? John P. HOEHN and Kwami ADANU, Michigan State Univ., East Lansing, MI. Discussant: Kim ROLLINS, Univ. Nevada Reno

Tuesday, 8:00 AM–12:00 PM: Paper Session 3: Paper Presentations Land Use: Recreation, Protection Measures, & Policy

Chair: John BERGSTROM, Univ of Georgia, Athens

Estimates and Welfare Analysis in a System of Correlated Count Outcomes. Joseph A. HERRIGES, Iowa State Univ, Ames

Spatial Limits of the TCM Revisited: Island Effects. John B. LOOMIS, Colorado State Univ, Fort Collins

Experiments on Response Rates for Stated Preference Surveys. Michael KAPLOWITZ, Frank LUPI, Michigan State Univ, E. Lansing

The Value of Information from Soil Surveys. Jerry FLETCHER, West Virginia University, Morgantown

An Analysis of Local Stakeholder Values for Tropical Protected Areas in Madagascar. Don DENNIS, Northern Research Station, USDA Forest Service, South Burlington, VT

Montana Challenge: Remaining the Last, Best Place for Fish and Wildlife in the Changing West. Cindy S. SWANSON, Rocky Mountain Research Station, USDA Forest Service, Missoula, MT

Building Wealth Through Ownership: Resident Owned Manufactured Housing Communities in New Hampshire. Kelly GIRAUD CULLEN, University of New Hampshire, Durham, NH

Tuesday 2:30 – 4:00: Panel Discussion: Publishing in Peer Reviewed Journals Today: A View from the Inside

Chair: Roger VON HEAFEN, North Carolina State Univ, Raleigh

Paul JAKUS, Utah State Univ, Logan, UT, Associate Editor/Editorial Council of J. Agricultural and Resource Economics, Society and Natural Resources, J. Environmental Economics and Management, Water Resources Research

John B. LOOMIS, Colorado State Univ, Fort Collin, Associate Editor/Editorial Council of Water Resources Research, and Journal of Agricultural and Resource Economics. Published 170+ Peer Reviewed Articles

Douglas MORRIS, Univ New Hampshire, Durham, Secretary/Treasurer of Northeastern Agricultural and Resource Economics Association, in charge of Oversight for Agricultural and Resource Economics Review

Randy ROSENBERGER, Oregon State Univ, Corvallis, Associate Editor of Journal of Leisure Research

Tuesday 4:15 – 5:45 PM: W2133 Business Meeting: Open to all W2133 Members

Chair: Michael Kaplowitz, Incoming W2133 President

Wednesday 8:00-12:00: Paper Session 4: Risk, Water, and Water Based Risk

Chair: Klause MOELTNER, Univ Nevada Reno

Alternative Futures Analysis for the Little Kanawha River Basin in West Virginia. Vishakha MASKEY, West Virginia University, Morgantown

Fire, Carbon, Timber, and Trees: Optimal Forest Management with Carbon Sequestration Credits and Endogenous Fire Risk. Brent SOHNGEN, Ohio State University, Columbus

Incorporating Protests Responses into Valuation Estimates: Willingness to Pay to Prevent Ecosystem Losses from Invasive Species and Wildfire. Kimberly ROLLINS, Univ Nevada Reno

Econometric Analysis of Environmental Water Values in Western U.S. Water Transactions. Bonnie COLBY, University of Arizona, Tuscon

Modeling Perceived Distributions of Mortality Risks from Arsenic Concentrations in Drinking Water. Paul JAKUS, Utah State Univ, Logan

In Pursuit of the Optimal Design: A Guide for Choice Experiment Practitioners[1]

William Breffle, PhD
Associate Professor of Natural Resource Economics
Michigan Technological University
School of Business and Economics
104 Academic Office Bldg.
1400 Townsend Dr
Houghton, MI 49931-1295
United States
wsbreffl@mtu.edu

March 1, 2008

[1] The author thanks Richard Bishop, Barry Solomon, Jennifer Thacher, an anonymous reviewer, and the participants of the 2008 W-2133 Western Regional Science Association Meeting in Kona, Hawaii for useful comments. The author acknowledges partial funding for portions of this paper from the U.S. National Oceanic and Atmospheric Administration. Any errors or misinterpretations of literature reviewed are the responsibility of the author.

Abstract

This paper is a primer for environmental economists who are developing a statistical design for a choice experiment application. While it will provide ideas for seasoned practitioners, it is especially for researchers using stated preference methods outside of the United States for the first time to value mounting ecological injuries and damages. The importance of a statistically efficient design is emphasized, but a multitude of other (sometimes competing) design factors are also discussed to demonstrate that there is no single "optimal" design. Using this paper as a guide and starting point, practitioners can avoid pitfalls early in the design stages, or at least be aware of their presence.

Key words: stated preference methods, choice questions, conjoint analysis, statistical design

1. INTRODUCTION

The conjoint method, also known as a "choice experiment," is frequently being used now to estimate public values for environmental amenities and reductions in disamenities, including ecological services. Over the past ten years, many environmental valuation practitioners have come to the conclusion that for many applications, choice experiments are preferable to the contingent valuation method (CVM). Hanley et al. (2001) does an excellent job comparing the two methods. See also Lienhoop and MacMillan (2007) for an updated discussion of CVM.

Both are survey-based value elicitation methods. In CVM, an environmental good is defined, the survey describes how it is to be provided, and then ultimately asks for willingness to pay for the good. Conceptually CVM is straightforward, but waning confidence in CVM stems from bidding approaches that can bias results, such as protest bids, yea-saying, and strategic bidding. In choice experiments, respondents are provided with a series of alternative states of the world, differing in terms of levels of attributes, and are asked to choose the most preferred. Some environmental applications include moose hunting in Canada (Boxall et al., 1996), preferences for different forest landscapes in the UK (Hanley et al., 1998), ecological services in Green Bay, Wisconsin (Breffle and Rowe, 2002), avoidance of nuclear meltdown risk (Itaoka et al., 2006), a Finnish conservation program (Li et al., 2004), and hydropower impacts and landscape improvements in Ireland (Lienhoop and MacMillan, 2007; Campbell et al., 2006).

The advantages of choice experiments are now widely known. Choice experiments generalize the CVM question and do a better job at measuring the marginal utilities and marginal rates of substitution among many attributes of a program, including money if desired (i.e., the marginal utility of money). This can reduce the cost of a study dramatically. Results are also more likely to show sensitivity to scope, as demonstrated by Forster and Mourato (1999). By avoiding the explicit elicitation of willingness to pay (WTP), one may avoid problematic responses, or at least make it more difficult for respondents to provide strategic responses (i.e., they are more likely to provide responses based on underlying preferences and a utility function).[2]

However, choice questions are no panacea, and can exponentially increase the cognitive burden on respondents. While many stated preference practitioners might agree that, all else constant, choice questions are preferred to contingent valuation, it is not the case that there is consensus in the literature about the "best" (i.e., optimal) design, and design needs to be assessed on a case-by-case basis. The optimal design might not be an efficient design, although much of the design literature possibly overemphasizes efficiency while ignoring other design concerns. Louviere (2006) notes that "adding realism is not a statistical design property" (p. 176), but realism is obviously an important aspect of a study.

This paper discusses the most important issues pertaining to the statistical design of a conjoint study. Maximizing efficiency is one of the most important topics, but not the only important topic. Design topics are discussed in general terms here, as are "rules of thumb". The content of this paper is based on a survey of existing design literature, particularly newer papers, and eight years of practical experience. While this paper provides a comprehensive starting point for understanding statistical design, other sources are available for additional advice and guidance, including Boxall et al. (1996), Carlsson and Martinson (2003), Ferrini and Scarpa (2007), Hanley et al. (1998), Holmes and Adamowicz (2002), Lusk and Norwood (2005), and Louviere et al. (2000).

[2] Theoretically, choice experiments can also be used to estimate willingness-to-accept for environmental degradation (i.e., conditions that are worse than the status quo). To my knowledge, Li et al. (2004) is the only study to do this.

2. APPROACHING THE DESIGN TASK

"Statistical design" is the organization of attributes and attribute levels among different choice sets. Given the number of observations (NOBS), the researcher must decide on (i) the number of attributes; (ii) the number of alternatives, and (iii) the number of choice sets. Then, one must also decide on the allocation of choice sets among survey versions. While a large sample can offset a poor design, that is not necessarily the best way to a approach the problem.

To begin, it is helpful to keep in mind three important concepts:

1. The efficiency of a design cannot be maximized without first knowing the real values of the parameters, which is a chicken-and-egg problem (Bunch et al., 1996). Finding the parameters is, after all, the point of the study. Further, the rewards of efficient designs are often unknowable (Ferrini and Scarpa, 2007).[3]

2. Many of the objectives that are important in developing the optimal design may directly conflict, which is to say that it is impossible to develop a design that is "all things to all people." As such, the researcher faces considerable subjectivity in weighing all of the important design elements to create the optimal design for a given application. This "optimal" design may not be best with regards to efficiency only or realism only.

3. It is essential that the researcher has already settled upon the functional form of the statistical model (i.e., the specification of the utility function), before creating the final design. Otherwise, one will not have controlled for the effects one is trying to estimate.

[3] Kanninen (2002) disagrees with this common belief. She shows that optimal designs can be constructed non-parametrically with updating, and using a single attribute, without knowledge of the true parameters. The downside is that the estimation precision will be diminished for the manipulator variable, which must be continuous.

Alpizar et al. (2001) suggest that the ex ante expected utility function might be approximated using other studies, expert judgments, pilot studies (see Campbell et al., 2006), and a sequential design strategy (basically a Bayesian approach where the demand parameters are successively updated as more information is added). See Scarpa et al. (2007) for an illustration of the efficiency gains derivable by a sequential updating of both categorical and continuous (e.g., price) variable levels using an iterative Bayes design procedure. A basic model can be run using choice experiment data from focus groups to generate starting values that can be used to help develop an efficient final design.

Bateman et al. (2006) conclude that larger sets of options are more realistic than pairs. They also show that pairs often lead to lexicographic decision making and unstable preferences. An example of unstable preferences is "loss aversion" (Holmes and Boyle, 2005). This happens when a respondent is first presented with a relatively high attribute level and has a lower marginal utility (and marginal WTP), and is subsequently presented with a lower level – something is taken away – and has a higher marginal utility. "Learning" can also cause unstable preferences (Holmes and Boyle, 2005). This happens when either the stochastic variance or the marginal utilities change with alternatives presented, as the respondent "learns" how to answer successive questions more accurately. However, when "triples" are used as opposed to pairs, for example, the number of choice set combinations increases exponentially.

With numerous alternatives in a choice set, Huber et al. (1982) show that not only can independence of irrelevant alternatives be violated, but that regularity conditions (specifically, within-subject preference reversals) can also be violated, limiting the range of applicability of discrete choice models. They find that when an alternative is added to a choice set and is similar to an existing alternative, the probability of choosing less-similar alternatives can actually increase. One explanation is that a consumer wants a product less if someone else can purchase a similar item.

A “full factorial” is a design that contains all possible combinations of the attributes and attribute levels. If the attributes and levels are numerous, or if the number of alternatives in the choice set is three or higher, the full factorial can be enormous. Therefore, a full factorial design is often unrealistic, when the number of choice sets per survey, the number of survey versions, and the sample size is typical [e.g., no more than four or five attributes (Alpizar et al., 2002)]. Garrod et al. (2002) is an example of a full factorial where dominant pairs (discussed later) are purged.

Four other design approaches have been used frequently (Lusk and Norwood, 2005):

1. Choice sets based on random drawing from the full factorial (a “fractional” design; see Itaoka et al., 2006)
2. A main-effects orthogonal design
3. A design based on maximizing some efficiency criteria
4. A design based on pragmatic reasons that is not concerned with statistical issues.

Some or most designs may be some combination of these four. Campbell et al. (2006) appropriately discuss potentially high survey costs as an important consideration in surveys of this type. At the other end of the spectrum, Louviere (2006) warns that it is not possible to fully learn how people make decisions without using a complete factorial, which for many researchers would be prohibitively expensive.

As a rule of thumb, Louviere et al. (2000) suggest a minimum of 32 alternatives when the full factorial is large, and NOBS = six or higher per alternative. Based on conjoint software simulations, Orme (1998) concludes that NOBS = n such that (n*t*a)/c should be greater than or equal to 500, where t = the number of choice tasks, a = the number of alternatives per task, and c = the largest number of level for one attribute, or the largest product if two-way interactions are used. If the population is segmented (e.g., there are multiple classes of individuals, where preferences differ across types by not within types), minimum NOBS = 200 per type.

Two examples of alternative combinations to be avoided are dominant pairs (where one alternative is clearly preferable over the rest) and unrealistic sequencing (e.g., a low level of an attribute with a high cost followed by a high level with a low cost). Campbell et al. (2006) employed an in-person survey staged in waves to adjust the monetary attribute in response to prior responses, which is not possible in a mail survey but may be possible using all other formats. But by extracting certain undesirable pairs, one moves away from perfect orthogonality, discussed in the next section. Designs with no dominant pairs are called "Pareto-optimal" by Wiley (2001), and compelling proof that these pairs should be avoided is presented: they are almost always identified as dominant (or dominated) by the respondent, so having them adds virtually no information, but can make parameters harder to identify. Campbell et al. (2006) found some random computer-generated landscapes to also be unrealistic.

3. WEIGHING EFFICIENCY AGAINST OTHER DESIGN FACTORS

The level of "efficiency" is the degree of precision of the model parameter estimates. With respect to estimating model parameters, there are really only two statistical aspects that are important: (i) unbiasedness/consistency; and (ii) efficiency. Assuming the model is unbiased, the researcher's main task is to maximize efficiency, all else constant. But all else may not be constant, and efficiency concerns may be at odds with problems such as the credibility of attributes or alternatives. Nevertheless, Louviere (2006), who first proposed discrete choice experiments in 1983 (Louviere and Woodworth, 1983), emphasizes that efficiency measures should be reported along with ability to generalize results "to verify claimed properties…and check against theoretical benchmarks" (p. 177).

There are four principles, first proposed by Huber and Zwerina (1996), that are relevant for obtaining efficiency (see also Alpizar et al., 2001):

1. orthogonality
2. level balance

3. minimal overlap
4. utility balance.

Their relationship with some types of efficiency are complex and even tenuous (Kanninen, 1993).

“Orthogonality” simply means that the attributes are purposefully uncorrelated in the design, which makes it easier to assess the effect of independent variables on dependent variables. “Level balance” means that there are the same number of attribute levels for every attribute. “Minimal overlap” means that attribute levels do not repeat. “Utility balance” means that alternatives in choice sets are close in utility space for respondents.

There are substantial issues and difficulties with each of these. “Caveats and severe assumptions must be made to derive optimal designs” p. 215 (Kanninen, 2002). Quantitative measures of efficiency can be employed, such as the D-optimal efficiency criteria. This is the most common design criterion, which seeks to maximize the determinant of the Fisher information matrix (Kanninen, 2002).

The D-statistic varies by what information the researcher has about the model’s demand parameters. There are essentially three categories (Ferrini and Scarpa, 2007):

a) no information, so an uninformed prior that all parameters = 0 is used
b) *a priori* information, based on a pretest or some other data
c) Bayesian information, where a priori information is progressively and cumulatively (i.e., hierarchically) added – for example, sequential survey waves. See Campbell et al. (2006) for a recent example.

In an earlier paper, Ferrini and Scarpa (2005) demonstrate that when the real data generating process is unknown, which is commonly the case in environmental applications, fractional designs from linear models produce less bias. However, with strong *a priori* information they advocate a D-optimal design.

Lusk and Norwood (2005) show through a Monte Carlo simulation that even with small sample, a D-optimal design for a linear-in-parameters multinomial logit model can "conveniently ignore" assumptions on the unknown parameter vector with little adverse effect (in other words, (b) and (c) above are ignored). In this case, the D-statistic is simply:

$$100/[N|(X'X)^{-1}|^{1/A}]$$

where N is the number of alternatives in a design, and A is the number of attributes * the number of attribute levels. X is the design matrix itself (rows = alternatives; columns = attributes). The value of the test statistic ranges from zero to 100, and is at a maximum when all four of the criteria are optimized. X'X is called the "information matrix" when the uninformed prior is used.

Ferrini and Scarpa (2007) and Campbell et al. (2006), in contrast, demonstrate that using a priori or Bayesian information can improve efficiency. Even with a minimal prior (e.g., a positive marginal utility of money), the efficiency gains obtained on a design with no information can be substantial. When some information about **β** is known, the information matrix takes its usual form, and the D-statistic is:

$$[|I(\boldsymbol{\beta})^{-1}|]^{1/k}$$

where k is the number of attributes. Ferrini and Scarpa reach the following conclusions, again based on a simulation of the real parameters:

a) D-efficiency is preferred when the specification and design are both correct.
b) D-efficiency with Bayesian information is preferred to "shifting" (defined below), which is in turn preferred to D-efficiency with *a priori* information when the specification is incorrect but the design is correct.

c) Shifting is preferred to any D-efficiency process when the specification is correct but the design is not[4] – this is the most common.

Two of these categories, (a) and (b), may not even be relevant: if we have the correct design, we do not need a design-creating process. No wonder category (c) is the most common! Therefore, one might conclude that using the shifting (or "cycling" method) is the safest approach.

Shifting is when one starts with an alternative, and then takes the values of all of the attribute levels and adds one (or another integer) to create the next alternative. You repeat the process until all possible alternatives using this process have been created. Put simply, the same value for any given attribute never appears twice, which makes the choice set as complex as possible (and the most burdensome on the respondent).[5] Both Breffle and Rowe (2002) and Itaoka et al. (2006) found estimation benefits in taking the opposite approach and starting the respondent off with questions that are as easy as possible during the learning process.

Kuhfeld et al. (1994) note that random drawing, while leading to independence of effects, does not maximize efficiency as measured using the D-efficiency statistic. They propose a three-step procedure to increase efficiency:

---

[4] The authors do not present compelling intuition on why this is the case – it is the result of a simulation experiment. "The more information is built into the design instead, the higher the degree of bias produced, *even under correct specification* [italics added for emphasis]" p. 16.

[5] With shifting one must take care to avoid choice sets with dominant alternatives. For example, suppose that there are three attributes, and all are "good", meaning that higher levels are preferred. Assuming three levels for each attribute, three alternatives in a choice set, and starting with {0,0,0}, adding one to each would give you {1,1,1}, which is preferred to the first; and adding two would give you {2,2,2}, which is preferred to both the first and the second. The only way to avoid this is to start with an alternative where all of the attribute levels (low, middle, and high) appear. {0,2,1}, {1,0,2}, and {2,1,0}, for example. Or {0,2,2,1}, {1,0,0,2} and {2,1,1,0} when there are four attributes. Note that the number of attributes must be at least as large as the number of alternatives in the choice set to avoid dominant alternatives. For example, consider a triple with two attributes. Three of the following combinations must appear: {0,0}, {0,1}, {1,0}, and {1,1} – and one of these alternatives will be dominating or dominated.

1. pull M alternatives randomly from the full factorial
2. compute D-efficiency
3. randomly replace alternatives one at a time and repeat until the test statistic changes by less than some small amount.

Just because one finds a replacement alternative that increases efficiency above some predetermined level, that does not mean that efficiency is optimized. There could be a combination that does better. That is, this process does not ensure global optimality (Ferrini and Scarpa, 2007), unless the process described above is repeated a large number of times, and the best matrix selected from each run. That is, the D-efficiency statistic likely will not automatically increase monotonically with this process; if the candidate matrix does not improve on the extant matrix, the candidate must be discarded and a better one sought.

Lew et al. (2006) propose another method to incorporate partial or incomplete information about model parameters into the design. This is “frequentist” model averaging, not a Bayesian method. The design is evaluated over a distribution of parameter values, with different designs weighted differently depending on the robustness to model and parameter uncertainty.

Orthogonality

Statistical programs can be used to generate choice sets with orthogonal attribute levels. However, prior to finalizing and administering the survey, the researcher must decide what variables are to be orthogonal. This ties in with the proper specification of the utility function.

In a “main-effects” design and model, interactions between variables (i.e., one variable multiplied by another variable) and heterogeneity among respondents is ignored. This is the simplest design, which is easy to develop and administer, and is relatively small.

Main effects dramatically reduces the number of choice sets. For example, a main-effects model with three alternatives and three levels yields 27 alternatives. Assuming 2-way interactions increases the number of alternatives to 243 (Lusk and Norwood, 2005). A full-factorial would require 24 versions with 10 questions each, and would not even take account of individual heterogeneity. However, assuming that attribute levels do not interact with one another (i.e., are uncorrelated with one another), or do not interact with some characteristic of the individual, may not only be naïve (Wiley, 2001), but may lead to biased estimates.

Louviere et al. (2000) notes that additive (i.e., main effects) models work well when the middle of utility space is of primary interest. That is to say, extrapolating to alternative combinations outside the relevant data range is not recommended. They go on to say that main effects often can ensure predictive accuracy even with an oversimplified specification. Bunch et al. (1996) concur. However, in a more recent article, Louviere (2006) warns that main-effects utility specifications may not be consistent with virtually all choice patterns in a choice experiment because interaction terms are typically present in the real utility function.

In most studies, main effects may explain 70% to 95% of the variance, and omission of 2-way interactions may not necessarily lead to bias (Dawes and Corrigan, 1974). Two-way interactions may account for another 15% of the explained variance; thus, three-way and higher interactions are typically unnecessary and make the design and the model much more complicated and more difficult to interpret. This is especially true when there are many attributes. Finally, with fractional designs, it is less possible to estimate interactions.

However, this has implications for the "status quo", where all levels remain at their current values, and the payment required is zero. The status quo is at the edge of utility and parameter space. There are good reasons to include the status quo:

1. It adds realism to the exercise

2. It is better than an opt-out alternative if the respondent does not like any of the alternatives presented.
3. It expands the range of attribute values, and thus expands utility space, which can be a positive as well as a negative.

Inclusion of the status quo obviously also increases the number of alternatives, and may have potentially large effects that inhibit creating an efficient design. Scarpa, Willis, and Acutt (2004) warn that if the status quo is included, then status-quo bias (which typically increases the probability that the status quo alternative is picked) should be checked. In their application, ignoring this bias can lower WTP values. Itaoka et al. (2006) found evidence of status-quo bias in their study as well.

Another consideration is that interactions may introduce extreme values for variables in the model (a small value multiplied by another small value, for example). Because of multicollinearity, it is advised that even when interactions are included, the squared components might be omitted (Louviere et al., 2000).

It may be unrealistic to assume that two attributes are uncorrelated (Louviere et al., 2000). For example, it would not be realistic to present the respondent with an alternative that protected a greater number of whales of a certain species than the total number of whales protected. The first variable level must be a lower bound on the second.

If many of the variables are dummy variables, the number of interactions is increased. If a preliminary evaluation can demonstrate that the marginal utilities on a set of dummies is essentially linear, then including a continuous variable will make the model more tractable. Another downside of including dummy variables is that the researcher is fundamentally unable to search the full design space in optimal construction, specifically, the spaces between the discrete values (Kanninen, 2002).

So, before a design is administered a correlation test across all variables, including interactions if they are important, should be done. A covariance matrix across the variables and their levels would also provide important insight.

Level balance

To maximize the power of the experiment, there must be the same number of levels for each attribute. Power is the ability to test the null hypothesis against multiple alternative hypotheses (Lusk and Norwood, 2005). To maximize efficiency, there needs to be level balance. Often, researchers use more levels for variables they deem to be more important (and thus, they seek more accurate estimates; e.g., the marginal utility of money). This is erroneous; the power of the test is constrained by the variable with the lowest number of levels. The smaller [the number of choice sets * the number of alternatives] is, the lower will be the power. Intuitively, level balance is analogous to the area of a parallelogram, which is maximized by maximizing and equalizing the length of each vector (i.e., a square) (Kanninen, 2002).

Minimal overlap

Minimal overlap essentially increases efficiency by increasing the cognitive burden upon the respondent. The general idea here is to make the choice sets as complicated as possible. For example, it requires that in a choice set, the same level of an attribute never appears twice.

Under the assumption that cognitive burden does not vary with complexity, minimal overlap is preferred. However, this is inconsistent with findings that the stochastic variance does in fact vary (Breffle and Rowe, 2002). Wiley (2001) states that, “when there are many attributes, even a single [alternative] may overwhelm respondents’ ability to process information” p.200.

Another concern is that inclusion of the status quo, which is common and often necessary when considering different programs (e.g., Holmes and Boyle, 2005), violates minimal overlap. In a referendum format, the status quo is repeated in every choice set.

To maximize the number of attribute comparisons, Bunch et al. (1996) recommend the shifting method, where the values of attributes are increased in a systematic way across alternatives as described above. The problems with shifting are that the researcher cannot test for scope (because many variables are changing at once), cannot include pairs with a minimum amount of cognitive burden (with only one attribute level changing), and it is inconsistent with the inclusion of the status quo (because that alternative always has the same attributes and levels).

Utility balance

Utility balance was first introduced by Bunch et al. (1996) and Huber and Zwerina (1996). Simply put, this aspect means that the researcher gets more information when the respondent has a more difficult time making choices. When two alternatives have a similar utility level for a respondent, it makes it difficult for him or her to choose, and we get a finer estimate of the marginal utility parameter. But this requires priors about what alternatives generate similar utility levels. If alternatives are too similar, the respondent will not be able to make a meaningful choice, which may only add noise (i.e., heteroskedasticity) to the model. Alternatives that are close in utility space may also lead to choice intransitivities, where preferences appear to be discontinuous (Huber et al., 1982; Wiley, 2001). The presence of many intransitivities may call into question the overall fitness of the design.

Kanninen's (2002) contention that "large contrasts within attributes provide more information about each parameter" (p. 223) appears to conflict with utility balance. Kanninen's optimality requires there to be only two levels per attribute, and also that the attribute values be as far apart as possible.

Finally, an "opt-out" alternative allows the respondent to avoid having to make forced choices when the decision is difficult (Champ et al., 2005). However, the opt-out alternative usually results in little useful information and may result in insignificant parameter estimates. Besides, the opt-out is really always available. A respondent can simply refuse to answer a question. The opt-out alternative can be misinterpreted as well. For example, Ruby et al. (1998) assume that when a "no-purchase" option is chosen, the respondent is indifferent between the two alternatives presented in a pair.

4. Concluding remarks

Developing an optimal design based on the four principles necessary to maximize statistical efficiency may not be as transparent as it initially seems. The effects of statistical design on the cohesion of respondent preferences and on realism must not be ignored (see DeShazo and Fermo for a discussion of the former). Not all of the relevant criteria in working toward an optimal design are statistically grounded.

The purpose of this paper is to take an inventory of many of the important issues in conjoint design without overburdening the reader with statistical proofs or conclusions from applications. It provides a framework for practitioners to organize their thoughts in beginning to develop a conjoint survey design that not only seeks to maximize statistical efficiency but is also sensible.

5. References

Alpizar, F., F. Carlsson, and P. Martinsson. 2001. Using Choice Experiments for Non-market Valuation. Working Papers in Economics no. 52, Department of Economics, Goteborg University, June.

Bateman, I., B. Day, G. Loomes, and R. Sugden. 2006. Ranking Versus Choice in the Elicitation of Preferences. Working paper. June 18.

Boxall, P.C., W.L. Adamowicz, J. Swait, M. Williams, and J.Louviere. 1996. A comparison of stated preference methods for environmental valuation. *Ecological Economics* 18(3):243-253.

Breffle, W. and R. Rowe. 2002. Comparing choice question formats for evaluating natural resource tradeoffs. *Land Economics* 78(2):298-314.

Bunch, D.S., J. Louviere, and D. Anderson. 1996. A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes. Working Paper UCD-GSM-WP#11.96. May.

Campbell, D., W.G. Hutchinson, and R. Scarpa. 2006. Using Discrete Choice Experiments to Derive Individual-Specific WTP Estimates for Landscape Improvements under Agri-Environmental Schemes: Evidence from the Rural Environment Protection Scheme in Ireland. The Fondazione Eni Enrico Mattei Note di Lavoro Series Index: Nota Di Lavoro 26.2006. Februrary.

Carlsson, F. and P. Martinsson. 2003. Design techniques for stated preference methods in health economics. *Health Economics* 12:281-294.

Champ, P., A. Alberini, and I. Correas. 2004. Using contingent valuation to value a noxious weeds control program: The effects of including an unsure response category. *Ecological Economics* 55:47-60.

Dawes, R. and B. Corrigan. 1974. Linear models in decision making. *Psychological Bulletin* 81:95-106.

DeShazo, J.R. and G. Fermo. 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 43(3):360-385.

Ferrini, S. and R. Scarpa. 2007. Designs with *a priori* information for nonmarket valuation with choice experiments: A Monte-Carlo study. *Journal of Environmental Economics and Management* 53:342-363.

Ferrini, S. and R. Scarpa. 2005. Experimental Designs for Environmental Valuation with Choice-Experiments: A Monte Carlo Investigation. Working Paper in Economics 8/05, Department of Economics, University of Waikato, New Zealand. December.

Forster, V. and Mourato, S. 1999. Elicitation Format and Part-Whole Bias: Do Contingent Valuation and Contingent Ranking Give the Same Result? CSERGE Working Paper GEC 99-117.

G.D. Garrod, R. Scarpa and K.G. Willis. 2002 Estimating benefits of traffic calming on through routes: a choice experiment approach. *Journal of Transportation Economics and Policy* 36:211-231.

Hanley, N., S. Mourato and R.E. Wright. 2001. Choice modeling approaches: A superior alternative for environmental valuation? *Journal of Economic Surveys* 15(3):435-462.

Hanley, N., R.E. Wright, and W. Adamowicz. 1998. Using choice experiments to value the environment – design issues, current experience and future prospects. *Environmental and Resource Economics* 11(3-4):413-428.

Holmes, T. and W. Adamowicz. 2002. Attribute Based Methods. In *A Primer on the Economic Valuation of the Environment*, eds. P. Champ, T. Brown and K. Boyle, Kluwer.

Holmes, T.P. and K.J. Boyle. 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. *Land Economics* 81(1):114-126.

Huber, J., J.W. Payne, and C. Puto. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research* 9(1):90-98.

Huber J. and K. Zwerina. 1996. The importance of utility balance in efficient choice designs," *Journal of Marketing Research*, 33 (August) 307-317.

Itaoka, K., A. Saito, A. Krupnik, W. Adamowicz, and T. Taniguchi. 2006. The effect of risk characteristics on the willingness to pay for mortality risk reductions from electric power generation. *Environmental and Resource Economics* 33:371-398.

Kanninen, B. 2002. Optimal design for multinomial choice experiments *Journal of Marketing Research* Vol. 39, 214-227.

Kanninen, B.J. 1993. Optimal experimental design for double-bounded dichotomous choice contingent valuation. *Land Economics* 69(2):138-146.

Kuhfeld, W.F., R.D. Tobias, and M. Garratt. 1994. Efficient experimental design with marketing research applications. *Journal of Marketing Research* 31:545-57.

Lew, D.K., D.F. Layton, and R.D. Rowe. 2006. Robust Experimental Designs for Economic Valuation Choice Experiments. Presented at the AERE session of the American Agricultural Economics Association Annual Meeting in Long Beach, CA, July 23-26.

Li, C., J. Kuuluvainen, E. Pouta, M. Rekola, and O. Tahvornen. 2004. Using choice experiments to value the Natura 2000 nature conservation programs in Finland. *Environmental and Resource Economics* 29:361-374.

Lienhoop, N. and D.C. MacMillan. 2007. Contingent valuation: Comparing participant performance in group-based approaches and personal interviews. *Environmental Values* 16:209-232.

Louviere, J.J. 2006. What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. *Environmental and Resource Economics* 34:173-188.

Louviere, J., D. Hensher, and J. Swait. 2000. Stated Choice Methods: Analysis and Application. Press Syndicate of the University of Cambridge, UK.

Louviere, J. and G. Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research* 20:350-367.

Lusk, J. and F.B. Norwood. 2005. Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics* 87(3):771-785.

Orme, B. 1998. Sample Size Issues for Conjoint Analysis Studies. Sawtooth Software Research Paper Series.

Ruby, M., F.R. Johnson, and K. Mathews. 1998. Just Say No: Opt-Out Alternatives and Angler's State Preferences. TER General Working Paper No. T-9801R. November.

Scarpa, R., D. Campbell, and W.G. Hutchinson. 2007. Benefit estimates for landscape improvements: sequential bayesian design and respondents' rationality in a choice experiment. *Land Economics* 83(4):617-634.

Scarpa, R., K.G. Willis, and M. Acutt. 2004. Monte Carlo Simulation Evidence on the Effect of the Status-quo in Choice Experiment Models for Water Supply. Presented at the Budapest conference of The European Association of Environmental and Resource Economists. June.

Wiley, J.B. 2001. Experimental Designs in Choice Experiments in *Recent Advances in Experimental Designs and Related Topics: Papers presented at the conference in honor of Professor Damaraju Raghavarao.* S. Altan, J. Singh, and D. Raghavarao, eds. New York: Nova Science Publishers.

Incorporating Random Coefficients and Alternative Specific Constants into Discrete Choice Models: Implications for In-Sample Fit and Welfare Estimates *

H. Allen Klaiber
Department of Agricultural and Resource Economics
North Carolina State University

Roger H. von Haefen
Department of Agricultural and Resource Economics
North Carolina State University

This Draft: March 2008

Abstract

In recent years, several innovative econometric methods have been employed in non-market valuation applications of discrete choice models. Two particularly attractive methods are random parameters (which introduce more plausible substitution patterns) and alternative specific constants (which control for unobserved attributes). In this paper, we investigate the properties of these methods along several dimensions. Across three recreation data sets, we consistently find large improvements in model fit arising from the inclusion of both methods; however, these gains often come concomitant with significant degradations in-sample trip predictions. We then show how poor in-sample predictions correlate with welfare estimates. Using econometric theory and Monte Carlo evidence, we illuminate why these perverse findings arise. Finally, we propose and empirically evaluate four 'second-best' modeling strategies that attempt to correct for the poor in-sample predictions we find in our applications.

Keywords: Discrete Choice, Recreation Demand, Revealed Preference, Stated Preference, Welfare Estimation, Alternative Specific Constants, Random Parameters

*Early draft that is currently under major revision. Comments welcome, but please do not cite. We thank George Parsons, Bill Breffle, Dan Phaneuf, Chris Timmins, and W2133 Annual Meeting participants for helpful comments.

Section I. Introduction

Discrete choice models have become one of the most frequently used modeling frameworks for recreation demand and locational equilibrium models (Murdock, 2006; Bayer and Timmins, 2007). Within the framework, two econometric innovations that applied researchers are using with increasing regularity are random coefficients (McFadden and Train, 2000) and the inclusion of alternative specific constants (Berry, 1994). Random coefficients are an attractive mechanism for relaxing the restrictive implications of the independence of irrelevant alternatives (IIA), thus introducing more plausible substitution patterns. Including a full set of alternative specific constants allows the analyst to control for unobserved attributes that may be correlated with observed attributes.

In applications of these modeling innovations to discrete choice models, researchers have found that they generate substantial and statistically significant improvements in fit (von Haefen and Phaneuf, 2008; Murdock, 2006). In an empirical investigation of three recreation data sets, we also find large gains in model fit. However, we also find that the models with alternative specific constants and random coefficients often fail to replicate the in-sample aggregate visitation patterns implied by the data. This empirical regularity generates important implications for the credibility of welfare analysis – why should one believe welfare measures derived from models that cannot replicate in-sample aggregate choice behavior?

Our goal in this paper is to shed light on the counterintuitive empirical regularity of improved statistical fit combined with poor in-sample prediction. We begin by documenting this phenomenon with three recreation data sets that have been used in previously published research. Two of the three applications combine revealed and stated preference (RP-SP) to identify all demand parameters (Adamowicz et al., 1997; Haener et al., 2001) as previously done by von Haefen and Phaneuf (2008). The other exploits only revealed preference (RP) data (Parsons et al., 1999) and uses a variation of the two-step estimator proposed by Berry, Levinsohn, and Pakes (2004) and used recently by Murdock (2006) in the recreation context. With all three data sets, we find the introduction of random coefficients and alternative specific constants (ASCs hereafter) substantially and significantly improves statistical fit as measured by the log-likelihood. We also find that in-sample trip predictions often (but not uniformly) deteriorate with these richer empirical specifications, and we document how these poor predictions correlate with welfare estimates for a range of policy scenarios.

We then explore why the poor predictions arise in practice. Here we use theoretical results from Gourieroux, Monfort, and Trognon (1984) about the properties of the linear exponential family of distributions as well as some Monte Carlo findings. The upshot of our discussion is that: 1) fixed coefficient logit models with a full set of ASCs will generate in-sample trip predictions for each alternative that *perfectly* match the data, and 2) random coefficient logit models with or without ASCs may not predict perfectly in-sample, but should generate reasonably close predictions if the analyst has correctly specified the underlying data generating process. An implication of this finding is that the poor in-sample predictions that we find in our three applications arise because of model misspecification. Thus, logit models with random coefficients and ASCs fit the data better than models without these econometric innovations, but they nevertheless fail to account for important features of the data.

We conclude by exploring a number of ‘second best’ strategies for dealing with poor in-sample predictions. These range from: 1) abandoning random coefficient specifications and using fixed coefficient models with ASCs that generate perfect in-sample predictions; 4) using less-efficient non-panel random coefficient models that, as we demonstrate, generate more plausible in-sample predictions; 3) using the Berry (1994) contraction mapping or maximum penalized likelihood (Montricher et al., 1975, Silverman, 1982; Huh and Sickles, 1994; Shonkwiler and Englin, 2005) with ASCs to force the in-sample predictions to match the data perfectly; and 4) conditioning on observed choice in the construction of welfare measures following von Haefen (2003). Our preliminary results suggest that each of these strategies is effective in terms of generating plausible in-sample predictions but they differ considerably in terms of their implications for statistical fit.

The paper proceeds as follows. The next section documents the performance of fixed and random coefficient logit models with and without a full set of ASCs with three recreation data sets. Section III explores the factors that give rise to the perverse empirical findings reported in the previous section using econometric theory and a set of Monte Carlo simulations. Section IV investigates a number of ‘second best’ empirical strategies that applied researchers may find attractive in future applications. We then conclude with some final observations and recommendations.

Section II. Nature of the Problem

We begin by illustrating the poor in-sample prediction problem that serves as the motivation for this research. To demonstrate that this problem is not an idiosyncratic feature associated with a single data set, we consider three recreation data sets that researchers have used in previously published studies. The first data set was first used by Adamowicz et al. (1997) and consists of both revealed preference (RP) and stated preference (SP) choice data for moose hunting in the Canadian province of Alberta. The RP data consists of seasonal moose hunting trips for 271 individuals to 14 wildlife management units (WMUs) throughout Alberta in 1993. The SP data consists of 16 choice experiments that were generated with a blocked orthogonal, main effects design. All eleven site attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. The second data set was first used by Haener et al. (2001) and also consists of combined RP-SP data for Canadian moose hunting. This data source, however, was collected in the neighboring province of Saskatchewan in 1994. The RP data consists of seasonal moose hunting trips for 532 individuals to 11 wildlife management zones (WMZs) throughout Saskatchewan. The SP data consists of 16 choice experiments that were generated with a blocked orthogonal, main effects design. All nine attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. As discussed in von Haefen and Phaneuf (2007), the fusion of RP and SP data is attractive in both data environments because the inclusion of a full set of ASCs confounds identification of the site attribute parameters given the relatively small number of sites in each application. For both data sets, we control for differences in scale across RP and SP data sources and use empirical specifications, estimation strategies, and welfare scenarios that match those used by von Haefen and Phaneuf (2008).

The third data set we consider looks at Mid-Atlantic beach visitation and was first used by Parsons et al. (1999). This data set consists of seasonal trip data to 62 ocean beaches in 1997 for 375 individuals. For each beach, we observe 14 site characteristic variables plus we construct individual-specific travel costs based on each recreator's home zip code. Because we use only RP data with this application, we use a two-step estimation strategy for those models that includes a full set of ASCs (Berry, Levinsohn, and Pakes, 2004; Murdock, 2006). For the results reported in Table 1, our two-step estimator differs from previous two-step estimators in the following way. Similar to Murdock, we use maximum likelihood techniques in the first step to estimate the travel cost parameter and a full set of ASCs that subsume all 14 site characteristics that do not vary over individuals (note: we do not include any demographic interactions in this model because preliminary testing suggested that they did not improve model fit). In contrast to Murdock, our first step estimator does not employ the Berry (1994) contraction mapping algorithm, an issue we return to in a later section. Thus, our first step estimator relies entirely on traditional maximum likelihood techniques, not the combination of maximum likelihood and Berry contraction mapping techniques that Murdock employs.[6] Our second-stage estimator is identical to Murdock's approach in that we regress the estimated ASCs from the first stage on the 14 site characteristics and a

[6] Using traditional maximum likelihood estimation techniques without the Berry contraction mapping is feasible in our application due to the relative small number of sites in the Mid-Atlantic data set. However, computational tractability requires the use of the Berry contraction mapping in random coefficient applications with many sites.

constant term. Importantly, this approach assumes that the unobserved site attributes are uncorrelated with observed site attributes.

Table I summarizes our findings.[7] All random coefficient models assume that the main effects for the site attributes (excluding travel cost) are normally distributed with no correlations. In on-going work, we are exploring truncated normal and latent class mixing distributions. Arrayed across columns 2-5 are results from four alternative specifications that differ in terms of the inclusion/exclusion of ASCs and random coefficients. In particular, column 2 contains results from models with neither ASCs nor random coefficients, column 3's results contain ASCs but no random coefficients, column 4's results contain random coefficients but no ASCs, and column 5's results contain both. Note that all random parameter specifications assume that all main effects for the various site attributes vary randomly across the population but are common for a given individual, so we refer to these specifications as 'panel' random coefficient specifications following Train (1998). Beginning first with the Alberta results, we note that relative to our baseline model without ASCs and random coefficients, the addition of these modeling innovations generates substantial improvements in fit. The largest gains seem to come from the addition of random coefficients that introduce correlations across an individual's multiple trips, although likelihood ratio tests suggest that ASCs also improve model fit significantly (p value < 0.0001).

To ascertain how well these models predict aggregate trip taking behavior for each site, we construct the following summary statistic for each model:

[7] Parameter estimates are available upon request.

(1)

$$\text{Percentage absolute prediction error} = 100\times\sum_{i=1}^{J} s_i^S \frac{\text{abs}(s_i^S - s_i^M)}{s_i^S} = 100\times\sum_{i=1}^{J}\text{abs}(s_i^S - s_i^M),$$

where s_i^S and s_i^M are the in-sample share of trips to site *i* and the model's prediction of the share of trips to site *i*, respectively, and *J* is the number of sites. The prediction error statistic can be interpreted as the share weighted in-sample prediction error for each site and thus can be used to rank order the models in terms of in-sample predictions that match the observed data. Intuitively, a model that can replicate aggregate trip predictions well for each site would generate a low prediction error value, whereas a model with poor in-sample aggregate predictions for each site would score a relatively high value. For the Alberta data, we see that the fixed coefficient specification with ASCs has the lowest prediction error statistic (effectively zero), whereas the random coefficient without ASCs has the highest. Interestingly, the substantially better fitting random coefficient with ASCs model has a prediction error statistic that is similar in magnitude to the more parsimonious fixed coefficient without ASCs specification.

Finally, it is interesting to see how these differences in fit and prediction play out in terms of welfare estimates. We consider two scenarios – a reduction in moose population at WMU #348 and an increase in moose population at WMU #344 – and calculate the partial equilibrium (i.e., ignoring changes in congestion) compensating surplus for both scenarios using the approach first suggested by Train (1998). In addition to point estimates and standard errors for the welfare measures, we also report the percentage in-sample prediction error for those sites directly affected by the different policies. Overprediction of the share of trips to these sites is likely to translate into larger welfare estimates, although variability in parameter estimates and the structure of substitution implied by the different models will also play a significant role. For the moose population reduction scenario, we find a range of point estimates from -$9.47 to -$25.00 with significant variation in these estimates' precision. The fixed coefficient with ASCs specification generates in-sample predictions for trips to WMU #348 that match the data well, whereas the other specifications overpredict trips to WMU #348 and generate larger (in absolute value) welfare estimates. For the moose population increase scenario, we find even larger variation in point estimates ($3.61 to $98.34) with significant variation in precision once again. In general, the smaller estimates correspond to specifications that underpredict the share of trips to WMU #344. Based on these results, we conclude that poor in-sample predictions play a significant role in explaining the variation of welfare point estimates in the Alberta data.

Similar results arise with Saskatchewan moose hunting data and the Mid-Atlantic beach data. With both data sets, adding ASCs and especially panel random coefficients improves statistical fit as measured by the log-likelihoods, but this improvement in fit does not necessarily generate lower prediction errors. The percentage absolute prediction errors for the fixed coefficient with ASCs models is once again near zero, but the percentage absolute prediction errors for the panel random coefficient models (with and without ASCs) are uniformly larger than the fixed coefficient models. For the Saskatchewan data, welfare point estimates and their precision vary significantly across the competing models. The variation in point estimates across the competing models seems to be correlated with the degree to which the models over- or underpredict trips to the affected sites. Finally, there appears to be considerably less variation in welfare point estimates for the Mid-Atlantic data, which may be explained by the fact that the alternative models seem to predict in-sample far better for the Mid-Atlantic data than the Alberta or Saskatchewan data.

In summary, the results in Table 1 suggest a somewhat counterintuitive result – including ASCs and especially random coefficients significantly improve overall statistical fit but do not generate in-sample trip predictions that match the observed data well. Welfare measures seem to be correlated with the degree of over- or underprediction implied by the different specifications, but other factors – parameter estimates, the structure of substitution implied by the models – certainly play a significant role. Overall, the results in Table 1 provide mixed evidence in favor of incorporating random coefficients and ASCs into discrete choice models, and cast doubt on the credibility of welfare estimates from models that predict in-sample poorly.

Section III. What explains these counterintuitive results?

In this section we use econometric theory and results from a Monte Carlo analysis to shed light on the counterintuitive results presented in the previous section. To motivate our main insight here, consider the log-likelihood function for a sample of N individuals each making separate choices from J alternatives:

$$\ln L(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{J} 1_{ij} X_{ij}\beta - \ln\left(\sum_{k=1}^{J}\exp(X_{ik}\beta)\right), \tag{2}$$

where 1_{ij} is an indicator function equal to 1 for individual *i's* chosen alternative and zero otherwise. The score condition associated with this log-likelihood is:

(3) $$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{N}\sum_{j=1}^{J} X_{ij}\left[1_{ij} - \Pr_i(j \mid \beta)\right] = 0\,,$$

where $\Pr_i(j \mid \beta)$ is the logit probability for individual *i* choosing the *j*th alternative. If a full set of ASCs are included, then

(4) $$X_{ij} = \begin{cases} 1 & \text{if } j \text{ chosen} \\ 0 & \text{otherwise} \end{cases}, \forall j\,,$$

and the score conditions associated with the ASCs can be written:

(5) $$\sum_{i=1}^{N}\left[1_{ik} - \Pr_i(k \mid \beta)\right] = 0 \quad \text{or} \quad \frac{1}{N}\sum_{i=1}^{N} 1_{ik} = \frac{1}{N}\sum_{i=1}^{N} \Pr_i(k \mid \beta), \forall k.$$

Equation 5 implies that fixed coefficient logit models with a full set of ASCs will generate in-sample predictions that match the data perfectly, a result that is consistent with our empirical findings in Table 1 and well known in the discrete choice literature (see, e.g., Ben-Akiva and Lerman, 1985).

As Gourieroux, Monfort, and Trognon (1984) have shown, the logit distribution falls within the broad class of distributions known as the linear exponential family of distributions. Other notable examples include the Poisson and normal distributions. What defines this family of distributions is that they are all mean-fitting distributions, implying that with the inclusion of ASCs, predictions from these distributions will match the data perfectly. A notable advantage of using linear exponential distributions in empirical work is that if the analyst has correctly specified the conditional expectation function of the distribution (i.e., its first moment), higher order misspecification will not lead to inconsistent parameter estimates (it will, however, bias standard error estimates, but this problem can be addressed if the analyst uses robust standard errors (White, 1981) instead of traditional standard errors). Thus, if the analyst specifies the first moment correctly, consistent parameter estimates will result. This makes the fixed coefficient logit model with ASCs appealing.

What is important to note, however, is that adding random coefficients to the logit distributions results in a mixture distribution that falls outside the linear exponential family. Random coefficient logit models, regardless of whether ASCs are included, will not necessarily generate in-sample predictions that match the data perfectly. This can be seen by looking at the score conditions for the simulated nonpanel random coefficient models logit model. The simulated likelihood function in this case is:

$$(6) \qquad L(\bar{\beta},\sigma)=\sum_{i=1}^{N}\sum_{j=1}^{J}1_{ij}\ln\left(\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j\mid\beta_i^r)\right)=\sum_{i=1}^{N}\sum_{j=1}^{J}1_{ij}\ln\left(\frac{1}{R}\sum_{r=1}^{R}\frac{\exp(X_{ij}\beta_i^r)}{\sum_{k=1}^{J}\exp(X_{ik}\beta_i^r)}\right),$$

where $\beta_i^r=\bar{\beta}+\sigma U_i^r$, $U_i^r\sim N(0,1)$, and the score condition is:

$$(7) \qquad \frac{\partial L(\bar{\beta},\sigma)}{\partial\beta_i^r}=\prod_{i=1}^{N}\left[\frac{\prod_{j=1}^{J}\left[X_{ij}\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j\mid\beta_i^r)(1-\Pr_i(j\mid\beta_i^r)\right]^{1_{ij}}}{\prod_{j=1}^{J}\left[\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j\mid\beta_i^r)\right]^{1_{ij}}}\right]=0\,.$$

With the inclusion of ASCs, this condition does not imply perfect in-sample predictions. Thus, some degree of imperfect in-sample prediction can be expected from random coefficient logit models, but the precise degree will vary across applications.

To assess how well in-sample predictions from estimated logit models will match the data, we conducted an extensive Monte Carlo analysis where we know the underlying data generating process for the simulated data. Knowing the true data generating process allowed us to ascertain the in-sample prediction performance of maximum likelihood estimators when model misspecification is absent. If the in-sample predictions generated from these correctly specified models match the observed data well, then we can conclude that poor in-sample predictions arise due to some form of model specification, and not due to an inherent property of the estimator.

For brevity, we only summarize the main conclusions of our Monte Carlo simulation here and leave for an appendix (to be written at a later date – apologies) the simulation details. Across a number of specifications, we consistently found that the in-sample predictions for panel and non-panel random coefficient models with and without alternative specific constants matched the simulated data very closely. Under none of our simulations did we find the degree of poor in-sample prediction that we observed with the Alberta, Saskatchewan, or Mid-Atlantic data – see Table 1. Based on these findings, we conclude that the poor predictions found in our three applications are a result of model misspecification.

The implications of the above discussion for how analysts should proceed are unclear. If the analyst estimates logit models with random coefficients and finds poor in-sample predictions, the obvious 'first best' solution would be to continue to search for empirical specifications that fit the data well and predict well in sample. In practice, however, finding empirical specifications that satisfy these two criteria will be computationally difficult, time-consuming, and in many cases infeasible. This suggests that 'second best' less demanding approaches that address these two concerns may be attractive alternatives to applied researchers. Perhaps the simplest second best approach would be to estimate a fixed coefficient logit model with ASCs where the in-sample aggregate predictions will match the data perfectly. One limitation with this approach is that it in practice employs models with substitution patterns that are consistent with the independence of irrelevant alternatives (IIA). These restrictive substitution patterns can be partially relaxed by using nested logit models, but the considerably more flexible substitution patterns that come with random coefficient models will not be realized.

Another second best approach involves estimating random coefficient models with ASCs using a contraction mapping (Berry, 1994) that iteratively solves for the ASC values by matching the aggregate model predictions with the data. This algorithm was first used in the industrial organization literature to estimate discrete choice models of product differentiation using aggregate market share data (Berry, Levinsohn, and Pakes, 1995), but Berry, Levinsohn, and Pakes (2004) apply the algorithm to a disaggregate data context. Both of these applications employed generalized method of moments estimation techniques, and it was not until Murdock (2006) that the algorithm was used within a maximum likelihood framework with random coefficients. What is interesting to note, however, is that the use of this algorithm within a maximum likelihood estimation procedure *will not* generate maximum likelihood estimates. For this to be the case, the random coefficient maximum likelihood estimates would have to generate in-sample predictions which match the data precisely, but we showed above that in general this will not be the case. Thus, the estimates that one recovers from using the Berry contraction mapping to estimate random coefficient models within the maximum likelihood estimates are akin to maximum penalized likelihood estimates that Shonkwiler and Englin (2005) and von Haefen and Phaneuf (2003) have previously used. The idea behind maximum penalized likelihood estimation is that one maximizes the likelihood subject to a function that penalizes the likelihood for some undesirable behavior. Random coefficient logit models with ASCs that are estimated within the maximum likelihood framework using the Berry contraction mapping are observationally equivalent to estimating random coefficient logit models with ASCs within the maximum penalized likelihood framework with an infinitely weighted penalty function for poor in-sample predictions. A limitation with this approach is that the asymptotic properties of maximum penalized likelihood estimators are not well understood, but it does directly address the poor in-sample prediction problem. Moreover, due to plateaus and non-concavities in the penalized likelihood function, the choice of starting values and search algorithms can strongly influence the derived estimates.

Two other second best approaches for dealing with poor in-sample predictions involve estimating non-panel random coefficient models with ASCs within the maximum likelihood framework or incorporating observed choice into the construction of welfare measures as suggested by von Haefen (2003). As we demonstrate in the next section, the former approaches sacrifice the efficiency gains (which may be substantial) from introducing correlations across an individual's multiple trips for improved (but not perfect) in-sample predictions. Moreover, it makes estimation more computationally intensive. The idea of incorporating observed choice into welfare measurement construction is attractive because it simulates the unobserved determinants of choice in a way that implies perfect prediction for every observation and then uses the model's implied structure of substitution to ascertain how behavior and welfare change with changes in price, quality, and income. The approach can be used with any set of model estimates, but it does require a somewhat more computationally intensive algorithm for calculating welfare estimates (see von Haefen (2003) for details).

In the next section, we compare the sensitivity of welfare estimates to the use of these four second best strategies that address poor in-sample predictions. Our discussion will focus on the Mid-Atlantic application where all welfare measures have been generated. In future revisions to this paper, we will fill in the missing estimates for the Alberta and Saskatchewan data to see how the approaches fair in these alternative data environments.

Section IV. Sensitivity of Welfare Measures to Alternative Second Best Strategies

The bottom third of Table 2 reports welfare estimates from the Mid-Atlantic beach data for two policy scenarios – lost beach width at all Delaware, Maryland, and Virginia (DE/MD/VA) beaches and the closing of all northern Delaware beaches. We report the log-likelihood values as well as the percentage absolute prediction error for all sites in the first two rows to give the reader a sense of the relative statistical fit and in-sample prediction performance of the competing specifications. We also report unconditional (Train, 199?) and conditional (von Haefen, 2003) welfare measures for both scenarios as well as the percentage prediction error at the sites directly affected by the policy for all specifications.

In general, the results reported at the bottom of Table 2 have a number of qualitative implications, although the reader should interpret these implications cautiously until they have been confirmed with the Alberta and Saskatchewan data. First, all of the second best strategies suggested in the previous section for dealing with poor in-sample predictions – using fixed coefficients and alternative specific constants (column 3), using the Berry contraction mapping (columns 8 and 9), and using non-panel random coefficient specifications with alternative specific constants (columns 6 and 8), as well as incorporating observed choice into welfare measures (the conditional welfare measures in all columns) are effective tools for mitigating this problem. Second, the use of non-panel random coefficients results in a significant loss of statistical fit (compare the log-likelihoods in columns 4 and 5, 6 and 7, and 8 and 9). Because the non-panel random coefficient specifications generate smaller prediction error relative to the panel random coefficient models, there is a significant tradeoff between statistical fit and good in-sample prediction when specifying the correlation structure of random coefficients. Third, using the Berry contraction mapping in estimation modestly degrades statistical fit (compare the log-likelihoods in columns 6 and 8 as well as 7 and 9), but it does improve in-sample predictions, especially when panel random coefficients are used.

In terms of welfare estimates, the results in Table 2 imply that there is little difference between the conditional and unconditional welfare across *all* specifications and scenarios. This result is not surprising because the in-sample trip predictions for the affected sites are generally small. For the lost beach width at DE/MD/VA beaches, we see most of the point estimates are clustered in the range of -$3.34 to -$11.76, although the estimates that are based on non-panel random coefficient models with ASCs (columns 6 and 8) are positive in sign. As suggested above, the non-panel random coefficient models fit the data far worse than the panel random coefficient models, and thus we doubt the reliability of these estimates which also have rather large standard errors. For the welfare scenario simulating the closing of northern Delaware beaches, we see a general convergence of estimates between -$11.92 and -$23.69. We believe this interval represents a plausible range of welfare estimates that should be sufficiently informative for policy purposes.

One could interpret the results from the Mid-Atlantic data as suggesting that the addition of ASCs and random coefficients has minor effects on policy inference. Indeed, the point estimates for the fixed coefficient model without ASCs are qualitatively similar to the mid-range values for the more complex specifications. Based on the incomplete set of results that are reported in Table 2 for the Alberta and Saskatchewan data, we doubt that this empirical finding will carry over to the other applications where prediction error is more extreme. However, one might conclude from the results presented in Table 2 that simple models that predict reasonably well in-sample might generate welfare estimates that are robust to the inclusion of alternative specific constants and random coefficients.

Section V. Conclusion

Our goal in this research has been threefold: 1) to document the somewhat counterintuitive in-sample prediction problems that arise with random coefficient logit models that include ASCs; 2) to explore the sources of these problems using economic theory and Monte Carlo analysis; and 3) to suggest and evaluate alternative, second best, strategies for dealing with the poor in-sample predictions that researchers might find attractive in future empirical work. Across three data sets, we document that the addition of ASCs and especially panel random coefficients generates significant improvements in statistical fit but do not uniformly improve model prediction. We also show how these poor predictions influence derived welfare estimates, with the degree of under- and overprediction at sites that are directly impacted by the policy being correlated with the magnitude of welfare estimates. We then argue that the fixed coefficient logit model falls within the larger family of linear exponential distributions, and thus the inclusion of a full set of ASCs will generate in-sample trip predictions for each site that match the data perfectly. The introduction of random coefficients, however, results in a mixture distribution that falls outside the linear exponential family and thus will not imply perfect in-sample predictions. Results from an extensive Monte Carlo analysis suggest that the poor in-sample predictions observed in our three applications are likely due to some form of misspecification. To account for these model shortcomings, the analyst may find attractive one of the second best strategies that we empirically evaluate for addressing poor in-sample predictions. Our preliminary empirical results with the Mid-Atlantic data suggest that all of these strategies are effective in controlling for poor in-sample predictions, but the use of non-panel random coefficients significantly degrades model fit and generates perverse signs for some of the policy scenarios. Otherwise, our results suggest that the other second best approaches imply qualitatively similar welfare estimates that fall within a narrow range.

Finally, it is worth stepping back and directly addressing the fundamental question that motivated this research: do random coefficients and alternative specific constants improve welfare analysis? With regard to random coefficients, we believe that the richer substitution patterns implied by random coefficients are quite attractive, but the poor in-sample predictions that often result from these models (especially panel random coefficient versions) need to be addressed in some way. If not, welfare estimates lack credibility. With regard to alternative specific constants, we believe that their ability to control for unobserved attributes that may generate endogenity concerns makes them extremely attractive. One limitation with their inclusion, however, is that one needs either an RP data set with many objects of choice (sites in recreation models, or neighborhoods in locational equilibrium models) or additional SP data to identify the part worths of the different site attributes. When these data are available, we believe that ASCs are an attractive modeling innovation.

References

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams, "Perceptions vs. objective measures of environmental quality in combined revealed and stated preference models of environmental valuation," *Journal of Environmental Economics and Management* 32 (1997): 65-84.

Bayer, P. and C. Timmins, "Estimating equilibrium models of sorting across locations," *Economic Journal* 117(2007): 353-374.

Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis*. MIT Press: Cambridge, 1985.

Berry, S. "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2): 242-262.

Berry, S., J. Levinsohn, and A. Pakes, "Automobile prices in equilibrium," *Econometrica* 63(1995): 841-890.

Berry, S., J. Levinsohn, and A. Pakes, "Differentiated products demand systems from a combination of micro and macro data: the new car market," *Journal of Political Economy* 112(2004): 68-105.

Gourieroux, C., A. Monfort, and A. Trognon, “Pseudo-maximum likelihood methods: application to Poisson models,” *Econometrica* 52(1984): 701-720.

Haener, M., P. Boxall, and W. Adamowicz, “Modeling recreation site choice: Do hypothetical choices reflect actual behavior?” *American Journal of Agricultural Economics* 83(2001): 629-642.

Huh, K., R. Sickles, “Estimation of the duration model by non-parametric maximum likelihood, maximum penalized likelihood and probability simulators,” *Review of Economics and Statistics* 76(1994): 683–694.

McFadden, D. and K. Train, “Mixed MNL of discrete response,” *Journal if Applied Econometrics*, 15(2000): 447-470.

Montricher, G., R. Tapia, J. Thompson. “Nonparametric maximum likelihood estimation of probability densities by penalty function methods,” *Annals of Statistics* 3(1975): 1329–1348.

Murdock, J., “Handling unobserved site characteristics in random utility models of recreation demand,” *Journal of Environmental Economics and Management* 51(2006): 1-25.

Parsons, G., D.M. Massey, and T. Tomasi, “Familiar and favorite sites in a random utility model of beach recreation,” *Marine Resource Economics*, 14(1999): 299-315.

Shonkwiler, J.S., and J. Englin, “Welfare losses due to livestock grazing on public lands: a count data systemwide treatment,” *American Journal of Agricultural Economics* 87(2005): 302-313.

Silverman, B. “On the estimation of a probability density function by the maximum penalized likelihood method,” *Annals of Statistics* 10(1982): 795–810.

Timmins, C. and J. Murdock, "A revealed preference approach to the measurement of congestion in travel cost models," *Journal of Environmental Economics and Management* forthcoming.

Train, K., *Discrete Choice with Simulation.* Cambridge University Press: Cambridge, 2003.

Train, K., "Recreation demand models with taste variation over people," *Land Economics*, 74:2(1998); 230-239.

von Haefen, R., "Incorporating observed choice into the construction of welfare measures from random utility models," *Journal of Environmental Economics and Management* 45:2(2003): 145-165.

von Haefen, R., D. Phaneuf, "Identifying demand parameters in the presence of unobservables: a combined revealed and stated preference approach," *Journal of Environmental Economics and Management* forthcoming.

Table 1 – Model Fits, In-Sample Predictions, and Compensating Surplus

Specifications				
Alternative Specific Constants?	No	Yes	No	Yes
Panel Random Parameters?	No	No	Yes	Yes
RP/SP Alberta moose hunting data from Adamowicz et al. (1997)				
Log-likelihood	-5,655.2	-5,376.7	-4,817.8	-4,521.6
Percentage improvement in log-likelihood	-	4.92%	14.8%	20.1%
Percentage absolute prediction error – all sites	30.0%	0.13%	45.6%	21.2%
CS for moose population reduction at WMU #348	-$14.11 (37.5)	-$9.47 (2.19)	-$25.00 (10.6)	-$20.91 (4.67)
Percentage prediction error at WMU #348	+10.8%	+0.12%	+16.3%	+6.60%

CS for moose population	$3.61	$98.34	$4.83	$73.02
increase at WMU #344	(2.50)	(31.0)	(3.19)	(23.3)
Percentage prediction error	-48.4%	+0.07%	-88.1%	
at WMU #344				+26.3%
RP/SP Saskatchewan moose hunting data				
from Haener et al. (2001)				
Log-likelihood	-7,655.3	-7,482.3	-6,658.2	-6,547.5
Percentage improvement in	-	2.26%	13.0%	14.5%
log-likelihood				
Percentage absolute	26.3%	0.17%	56.8%	
prediction error – all sites				33.6%
CS for moose population	-$18.55	-$14.69	-$81.47	-$61.62
reduction at WMZ #59	(7.52)	(2.99)	(11.5)	(9.77)
Percentage prediction error			+82.9%	+30.7%
at WMZ #59	+31.5%	-0.02%		
CS for moose population	$27.54	$150.50	$22.30	$74.59
increase at WMZ #66	(4.10)	(36.9)	(3.36)	(14.3)
Percentage prediction error			-31.3%	+13.2%
at WMZ #66	-39.2%	+0.20%		
RP Mid-Atlantic beach data				
from Parsons et al. (1999)				
Log-likelihood	-13,160.2	-12,981.8	-11,015.8	-10,869.2
Percentage improvement in	-	1.36%	16.3%	17.4%
log-likelihood				
Percentage absolute	13.3%	<0.01%	27.0%	31.4%
prediction error – all sites				
CS for lost beach width at	-$6.44	-$4.89	-$5.64	-$7.57
DE/MD/VA beaches	(1.16)	(4.62)	(1.41)	(3.26)
Percentage prediction error	-2.22%	<0.01%	-12.1%	-1.75%
at DE/MD/VA beaches				
CS for northern DE beach	-$19.56	-$21.88	-$14.97	-$16.83
closings	(0.64)	(3.94)	(1.45)	(3.24)
Percentage prediction error	-5.04%	<0.01%	-6.01%	-8.28%
at northern DE beaches				

Robust standard errors in parentheses. All welfare estimates are per trip.

Table 2 – Alternative Strategies

Specifications								
Alternative Specific Constants?	No	Yes	No	No	Yes	Yes	Yes	Yes
Berry Contraction Mapping?	-	No	-	-	No	No	Yes	Yes
Random Parameters?	No	No	Panel	Non-Panel	Non-Panel	Panel	Non-Panel	Panel
RP/SP Alberta moose hunting data from Adamowicz et al. (1997)								
Log-likelihood	-5,655.2	-5,376.7	-4,817.8	-5,626.7	-5,368.3	-4,521.6		
Percentage absolute prediction error – all sites	30.0%	0.13%	45.6%	32.3%	0.14%	21.2%		
Unconditional CS for moose population reduction at WMU #348	-$14.11	-$9.47	-$25.00	-$23.09	-$9.56	-$20.91		
	(37.5)	(2.19)	(10.6)	(6.08)	(3.25)	(4.67)		
Conditional CS for moose population reduction at WMU #348	-$13.29	-$9.90	-$22.04	-$36.53	-$10.21	-17.01		
	(2.86)	(1.54)	(3.91)	(7.85)	(7.85)	(1.88)		
Percent. predict. error at WMU #348	+10.8%	+0.12%	+16.3%	+25.4%	+0.04%	+6.60%		
Unconditional CS for moose population increase at WMU #344	$3.61	$98.34	$4.83	$2.76	86.81	$73.02		
	(2.50)	(31.0)	(3.19)	(1.70)	(27.4)	(23.3)		
Conditional CS for moose population increase at WMU #344	$5.97	$98.00	$2.99	-$1.70	85.33	61.02		
	(7.17)	(29.8)	(2.37)	(1.62)	(23.5)	(20.6)		
Percent. predict. error at WMU #344	-48.4%	+0.07%	-88.1%	-45.2%	-0.20%	+26.3%		
RP/SP Saskatchewan moose hunting data from Haener et al. (2001)								
Log-likelihood	-7,655.3	-7,482.3	-6,658.2	-7,587.2	-7,472.9	-6,547.5		
Percentage absolute prediction error – all sites	26.3%	0.17%	56.8%	32.5%	4.84%	14.47%		
Unconditional CS for moose population reduction at WMZ #59	-$18.55	-$14.69	-$81.47	-$41.39	-$27.53	-$61.62		
	(7.52)	(2.99)	(11.5)	(16.3)	(12.1)	(9.77)		
Conditional CS for moose population reduction at WMZ #59	-$12.64	-$14.68	-$40.23	-$34.71	-$30.67	-$42.72		
	(5.31)	(2.65)	(4.57)	(12.4)	(12.1)	(4.58)		

Percent. predict. error at WMZ #59	+31.5%	-0.02%	+82.9%	+37.0%	+2.34%	+30.7%		
Unconditional CS for moose population increase at WMZ #66	$27.54	$150.50	$22.30	$25.14	$85.27	$74.59		
	(4.10)	(36.9)	(3.36)	(8.41)	(19.0)	(14.3)		
Conditional CS for moose population increase at WMZ #66	$32.82	$150.62	$26.57	$24.50	$79.56	$80.27		
	(4.18)	(36.3)	(3.68)	(8.85)	(17.0)	(14.7)		
Percent. predict. error at WMZ #66	-39.2%	+0.20%	-31.3%	-39.6%	+3.22%	+13.2%		
RP Mid-Atlantic beach data from Parsons et al. (1999)								
Log-likelihood	-13,160.2	-12,981.8	-11,015.8	-13,021.4	-12,856.5	-10,869.2	-12,874.5	-10,962.9
Percentage absolute prediction error – all sites	13.3%	<0.01%	21.8%	11.0%	+1.81%	+25.9%	<0.01%	<0.01%
Unconditional CS for lost beach width at DE/MD/VA beaches	-$6.44	-$4.89	-$5.64	-$3.34	$7.51	-$7.57	$1.83	-$11.76
	(1.16)	(4.62)	(1.41)	(0.60)	(7.72)	(3.26)	(3.96)	(4.27)
Conditional CS for lost beach width at DE/MD/VA beaches	-$6.58	-$4.89	-$7.15	-$3.59	$6.95	-$7.35	$1.78	-$11.53
	(1.18)	(4.27)	(1.11)	(0.70)	(7.16)	(1.79)	(3.70)	(1.95)
Percent. predict. error at DE/MD/VA beaches	-2.22%	<0.01%	-11.8%	-2.36%	+0.35%	-2.50%	<0.01%	<0.01%
Unconditional CS for northern DE beach closings	-$19.56	-$21.88	-$14.97	-$12.27	-$11.98	-$16.83	-$11.92	-$22.23
	(0.64)	(3.94)	(1.45)	(0.58)	(4.55)	(3.24)	(2.46)	(3.84)
Conditional CS for northern DE beach closings	-$20.75	-$22.04	-$16.58	-$13.54	-$13.51	-$19.34	-$13.27	-$23.69
	(0.69)	(2.34)	(2.09)	(0.69)	(1.76)	(1.88)	(1.09)	(2.54)
Percent. predict. error at northern DE beaches	-5.04%	<0.01%	-6.01%	-1.20%	-0.49%	-8.47%	<0.01%	<0.01%

Robust standard errors in parentheses. All welfare estimates are per trip.

A Hybrid Individual-Zonal Travel Cost Model for Estimating the Consumer Surplus of Golfing in Colorado

John Loomis, Omer Tadjion, Philip Watson, Josh Wilson, Steve Davies and Dawn Thilmany. Dept. of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO 80523-1172.

Corresponding author: John Loomis, Dept. of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO 80523-1172
Email: jloomis@lamar.colostate.edu
Phone: (970) 491-2485 FAX: (970) 491-2067

October 23, 2007

Acknowledgements: Funding for this study provided by Environmental Institute of Golf, and the Colorado Agricultural Experiment Station project W2133. Paul Huszar kindly gave us many insightful comments that could only be provided by someone that is both an avid golfer and economist. Of course, any remaining errors are the responsibility of the authors.

A Hybrid Individual-Zonal Travel Cost Model for Estimating the Consumer Surplus of Golfing in Colorado

Abstract

Using a survey of Colorado golfers and a relatively novel hybrid individual observation and zonal travel cost model, we find the demand for golf is quite price inelastic with respect to transportation costs (-.28) and green fees (-.14). The typical golfer spends $80 on transportation and $49 on green fees/carts. The price inelastic demands translate into a consumer surplus of $28.80 per round of golf at Colorado golf courses. The annual net economic value to golfers in Colorado for the 7.8 million rounds of golf is $224.64 million. We find a quadratic relationship between age and golf demand, such that retirement age golfers take about 30% more trips than middle age golfers, a trend that should bode well for future demand for golf as Baby Boomers age and increase their annual number of trips.

Keywords: consumer surplus, demand, green fees, price elasticity, travel cost method

Introduction

Golfing in the United States is a popular recreation activity with an estimated 38 million people that golf (National Golf Foundation, 2004). The high levels of spending by golfers generates $30 billion a year, a significant effect on the economy (National Golf Foundation, 1998). The nearly 15,000 golf courses themselves are both major land uses in many urban areas and major water users in the western U.S. Many residential neighborhoods built around golf courses feature the course as a central amenity and market themselves accordingly. Most destinations that consider themselves resorts, also prominently feature golf courses as one of their attractions.

However, as Correia et al (2007) recently noted, golf is an under-researched recreational activity. Almost nothing is known about the economic benefits (i.e., consumer surplus) that the golfers themselves receive. At municipal courses, which are often priced below private courses, there is likely to be a significant consumer surplus realized by golfers. From the standpoint of economic efficiency, the loss to society from conversion of old golf courses to other land uses would be the lost consumer and producer surplus, not the spending (which would be reallocated to other leisure activities).

Few Past Studies on Economic Demand for Golfing

While recreation activities like hunting, fishing, hiking and camping have had dozens of economic valuation studies, there have been very few studies on the demand for and economic benefits of golf. The first published study was by Milam and Pasour in 1970. They used golfer data from North Carolina to estimate the demand for golfing, and found it to be slightly price elastic for most categories of golfers. No consumer surplus or benefits are estimated, and given their unusual functional form, it would be difficult to calculate one from the data given in the paper.

Pricing practices at golf courses have been investigated by two authors. Shmanske (2001) estimates an aggregate course demand for rounds of golf by using course data from 47 San Francisco Bay area municipal golf courses. Shmanske (2001) found a very price inelastic demand. Mulligan (2001) investigates whether the use of membership fees as a way to price private golf courses like a club good are efficient or not. Mulligan concludes that when members' opportunity costs of time and congestion are considered that a "members only" course may in fact be efficient.

Recently, Correia et al (2007) estimated a repeat choice mixed logit model to investigate which golf course attributes and other destination characteristics influence return trips by golfers to the Algarve region in southern Portugal. Unfortunately the authors do not report any estimate of economic welfare nor is it possible to calculate one from their model.

Thus the purpose of this paper is to fill this void in the literature regarding the consumer surplus of golfing. We utilize a novel hybrid travel cost model to estimate the demand for and consumer surplus for golfing at public and private golf courses in Colorado. In the next section we develop the hybrid travel cost model, then we present the data, results and conclusion.

The Hybrid Individual Observation Zonal Travel Cost Demand Model

The Travel Cost Method (TCM) has evolved over the almost fifty years since Clawson first proposed the model in 1959 (Clawson and Knetsch, 1966).). The dependent variable in the first TCM models was the number of trips coming from a zone, divided by the population of that zone, i.e., trips per capita. The zones were originally concentric circles around the site, but now it is often counties or zip codes around the site to make use of computerized demographic data and to increase the spatial resolution of the unit of observations (Loomis and Walsh, 1997). This zonal approach is quite useful when applying the TCM to situations where the visitor data is from secondary sources such as recreation permits or fee receipts. The model is also quite useful for sites where each individual visitor takes just one trip per year (or there is data only on the most recent trip).

The limitation of this zonal model include statistical inefficiency due to the fact that aggregating the individual observations by zone averages out some of the information available in the individual data (Brown and Nawas, 1973). In Brown and Nawas' empirical example, it would require an aggregate sample 12 times as large to yield equivalent precision in the individual observation coefficients. In addition, if one did have some limited data on the demographic variables of visitors, only the zonal average of these could be used on the right hand side. In this situation, only cross zone variations in demographic variables would be reflected. Most importantly, in zonal models the zone average travel cost and zone average travel time cannot be separately included because they are nearly perfectly correlated (Brown and Nawas, 1973). As has been demonstrated by Cesario and Knetsch (1970), omission of travel time as a variable will bias the travel cost coefficient, and hence the consumer surplus. Thus economists have often had to monetize travel time by its opportunity cost and combine with travel cost to create a full price variable. The difficulty here is that consumer surplus estimates may be quite sensitive to the fraction of the wage rate used to value the opportunity cost of time (Bishop and Heberlein, 1979).

These limitations of the zonal TCM and the availability of individual visitor survey data gave rise to the individual observation TCM (Brown and Nawas, 1973; Gum and Martin, 1975). In this model the dependent variable is the number of trips an individual visitor makes each season or year and each visitor is a unit of observation. This model has become one of the dominant forms of TCM. Currently, these models are more efficiently estimated using count data estimators to account for the fact that individual trips to a site are non-negative integers (Creel and Loomis, 1990; Hellerstein, 1992).

Brown, et al. 1983 first suggested a hybrid individual-zonal TCM. This model uses the individual visitor as a unit of observation and allows maintaining the individual travel cost, travel time and demographic variables on the right hand side. However, the dependent variable is calculated by dividing that individual's visits by his/her share of the "zones" population to calculate trips per capita. The share of the population depends on how many visitors came from that zone of origin. For example, if there are three visitors from County A, then each visitor would be allocated one-third of the county population. Their trips to the site would then be divided by one-third the county population to arrive at their trips per capita. Zones with more visitors would have smaller shares of the population allocated to each visitor. Thus a zone with ten visitors, each would get one-tenth of the population. Thus if there are two zones of equal population, and one has two visitors and one has ten visitors, then the trips per capita would be higher for the zone with ten visitors than the one with two visitors. This approach reflects the fact that closer zones not only are likely to have visitors with higher trip frequency per visitor but also higher participation rates from their population. This form of the dependent variable works well if visitors only take one trip per year or the analyst only has data indicating that a visitor from that zip code or county made at least one trip to the site, but not how many individual trips he or she made over the course of the year.

Thus the hybrid individual-zonal travel cost model uses trips per capita as the dependent variable like the zonal model, but also includes individual level data on the right hand side. Therefore, travel time can be included as a separate variable, avoiding the necessity of having to assign a particular fraction of the wage rate to monetize the opportunity cost of time to combine with the travel cost variable. In addition, individual level demographics can be included. As will be apparent in the following discussion of the data, the structure of this model is quite advantageous given the structure of our data. The essential structure of this hybrid TCM is illustrated in equation (1):

(1): $IndTrips_{izj}/(Pop_z/V_{zj}) = Bo - B_1(TC_{zj}) - B_2(\text{Cost of Travel Time}_{zj}) + B_3(Age_i) + B_4(Gender_i) + B_5(Fees_{ij}) + B_6\ Income_i + B_7(GCMountains)$

Where: $IndTrips_{izj}$ is the number of trips by golfer i living in zip code z to golf course j. Pop_z is the population of zip code z, and V_{zj} is the number of golfers from zip code z that visit course j. Travel $Cost_{zj}$ is the calculated round trip distance from the golfer's residence zip code (z) to the golf course j, multiplied by the variable automobile cost per mile. Cost of Travel Time is the full wage rate times the calculated travel time from the golfer's residence zip code to the golf course. Since travel time is entered as a separate variable (rather than combining it with Travel Cost), whether we use the full wage rate or some fraction will not affect the travel cost coefficient and hence the consumer surplus.

Age_i is golfer's age in years

$Gender_i$ is whether the golfer is male or female

$Fees_{ij}$ is the cost of green fees and any cart fees paid by golfer i at course j.

GC Mountains is a dummy variable equal to 1 if the golf course is in the scenic Rocky Mountains and =0 if it is on the Front Range or eastern plains of Colorado.

$Income_i$ is the golfer's income.

Colorado Golf Course Data

This study used data gathered in a 2003 survey of golfers at Colorado golf courses. The primary aim of the survey was to investigate the economic contribution of golf to the Colorado economy, particularly in relation to golf course water use (Wilson, 2005). A total of 635 golfers were interviewed at nineteen golf courses throughout Colorado. Among them, eight are located in the Rocky Mountains, seven in the Denver metro- area and the remaining four are distributed in the northern and southern Front Range cities surrounding Denver.

Construction of Variables from the Survey Data

Visitation rate (INDTRIPCAP) is the dependent variable. Since the survey did not record the annual number of trips, the available data simply indicates that an interviewed golfer was from a particular zip code. Essentially, the number of trips per golfer is treated as one trip. This is then divided by the golfer's share of his or her zip code's population, where share is based on the reciprocal of the number of golfers from that zip code going to that particular golf course. The population of each zone (zip code) is provided by US Census Bureau (US Census Bureau, 2000). We scaled each zip code's population to thousands in order to minimize the number of leading zero decimal points.

One of the assumptions of the TCM to interpret the travel cost as part of the price of the trip, is that visitors incur the travel costs solely for the purpose of visiting the recreation site. To conform with this assumption, we limited our analysis to golfers who were residents of Colorado. This resulted in the maximum distance of slightly more than 200 miles, and mean and median distances of 31 miles and 11 miles, respectively. These distances also make plausible that golfers drove to the courses.

Thus, the price variable, travel cost, was constructed by calculating the round trip driving distance between golfer residence zip codes and the golf courses zip codes. This distance is computed using the zip code Distance Calculator (Imacination Sofware, 2002). Then the mileage is converted to round trip transportation cost, computed by using the standard average variable cost of 13 cent per mile for operating an automobile in the US (US Department of Transportation, 2003).

To calculate travel time from the round trip distance we use an average speed of 32 miles per hour from the National Highway Institute (1995) for golfers living no farther than 30 miles from the golf course, as these golfers would generally be traveling surface streets. For golfers traveling more than 30 miles, the travel speed beyond the first 30 miles is assumed to be 64 miles per hour due to these travelers using freeways, interstates and highways. The value of the travel time variable is then calculated as the product of the round trip total time spent and of the golfer's wage rate. It is treated as a separate variable to control for travel time, and not added into the travel cost.

The survey did report what the golfer paid for green fees and whether the golfer paid for a golf cart. At busy time periods, some golf courses require carts to speed up play, so it would be considered an exogenous cost like green fees. Mountain golf courses usually require golf carts as well.

The model also contains several demographic and qualitative variables such the golfer's gender (1 if male and 0 if female) and age (in years). A dummy variable is included for whether or not the golf course is located in the mountains and is equal to 1 if the golf course is in the Rocky Mountains, 0 if not). Golfer income was also included but it was later dropped due to high item non response and lack of statistical significance.

We performed a natural log transformation of the dependent variable for several reasons. First this resulting semi-log functional form mimics the commonly used functional form associated with count data models (e.g., Poisson and negative binomial). This functional form also simplifies the consumer surplus calculation, as consumer surplus per trip is just the reciprocal of the travel cost coefficient (Creel and Loomis, 1990). This also simplifies calculation of the confidence interval on the consumer surplus estimate. Finally, the natural log of the dependent variable allows for non-linearity in the demand function.

Results

Regression Results

The estimated individual observation per capita demand curve is reported in Table 1. All variables are statistically significant at the 5% level or better, and the F statistic is significant at the 1% level. While the R square is only 18%, Brown and Nawas (1973) note this is common with individual level data. As expected by economic theory, the coefficients on travel cost and travel time, as well as green and cart fees are all negative. The relationship between age and number of golf trips is quadratic. Younger golfers (e.g., 20 year olds) take slightly more trips (3.60) than middle age golfers, who at the sample mean age of 43 years, take 3.18 trips. However, the number of trips rises rapidly at retirement age, increasing to 4.38 trips at age 65 and 5 trips at age 70, representing 30% and 50% increases in frequency relative to the sample average age.

Table 1. Regression Results for Colorado Golf Demand Equation

Dependent Variable: Log of Individual Trips per Capita
Observations: 544

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	-0.771077	0.36313	-2.1233	0.0342
Travel Cost	-0.034725	0.01042	-3.3296	0.0009
Cost of Travel Time	-0.003255	0.00146	-2.2181	0.0270
Age	-0.033382	0.01656	-2.0154	0.0444
Age Squared	0.000444	0.00019	2.2979	0.0220
GC Mountains	0.449024	0.13517	3.3218	0.0010
Gender	-0.293511	0.13804	-2.1262	0.0339
Greens & Cart Fees	-0.002893	0.00126	-2.2914	0.0223
R-squared	0.1863	Mean dependent var		-2.0347
Adjusted R-squared	0.1757	S.D. dependent var		1.2339
S.E. of regression	1.1203			
F-statistic	17.539	Prob (F-statistic)		0.000000

Price Elasticities

Using the travel cost and green/cart fees coefficients in Table 1, along with the means of the respective coefficients, we calculate the price elasticities for these two costs. The price elasticity for travel costs is -0.28, while for green fee/cart costs it is half that at -0.14. Both are quite price inelastic, although the responsiveness to travel cost is twice that of green fees.

Our finding of price inelastic demand is consistent with Shmanske’s (2001) finding of price inelastic demand. However, both the present study and Shmanske’s (2001) price inelastic demand stands in contrast to the Milam and Pasour (1970) who found price elastic demand. It is possible that in the intervening 30 years since Milam and Pasour’s (1970) study that golf is less price elastic than in 1970, possibly due to the large number of municipal courses available at relatively low green fees.

As Shmanske (2001) notes, given the finding of price inelastic demand, if golf courses wished to increase revenue, they certainly do so by modest increases in green fees.

Consumer Surplus

Using the reciprocal of the travel cost coefficient, the consumer surplus per day of golf is $28.80 (with a 90% confidence interval of $19 to $57). With an estimated 7.8 million rounds of golf played in Colorado in 2003 (Davies, et al. 2004), this translates into a net economic value of golfing of $224.64 million annually.

The average consumer surplus can be compared to the average round trip travel cost of $8 at the time of the survey. However, the average green fee paid was $41. The average cart fee was just $8, although the median cart fee was zero as the majority of golfers did not rent a cart. Using the sum of these three cost elements and the consumer surplus, the gross willingness to pay can be calculated at $85.80 per day of golfing. Of the $85.80, the average Colorado golfer spent $8 to travel to and from the golf course and spent $49 on green fees and rental carts. That leaves a consumer surplus of $28.80 received by the golfer his/herself.

As noted in the literature review, there are no other estimates of consumer surplus of golfing to compare our estimates to. There are however, at least two rough comparisons possible, one based on a similar outdoor physical activity (hiking) and the other based on an activity with a similar pricing structure (downhill skiing). Playing 18 holes of golf involves walking approximately 5 miles at most courses. Given that vast majority of our golfers surveyed did not use a cart, and golfers essentially “hiked” 5 miles while playing golf. The most recent estimates of consumer surplus for hiking in Colorado is $39 (Loomis, 2005), somewhat higher than our estimate of golfing. In terms of pricing, downhill skiing requires purchase of a lift ticket which is not only analogous to a green fee/cart fee in terms of extracting consumer surplus from the user in exchange for access but similar in magnitude to lift ticket prices. The few available downhill skiing studies have a consumer surplus of $33.50, about 15% higher than the golfing consumer surplus.

Conclusion

This study adapts a hybrid individual observation per capita travel cost demand model to estimate the price elasticity and consumer surplus for golfing in Colorado. The results indicate that the demand for golfing in Colorado is quite price inelastic with respect to both transportation costs and green fees in 2003. The price inelastic nature of travel cost bodes well for golf courses in the face of the rather steep increases in gasoline prices in the last two years. Golf courses interested in increasing their revenues could also exploit their price inelastic demand and raise green fees somewhat. The quadratic relationship between age and number of rounds of golf suggests that the demand for golf should rise over time as the proportion of the population of retirement age (which takes 30% more trips) increases and the proportion of the population in middle age (which has the lowest estimated number of rounds) decreases.

The gross willingness to pay for a round of golf at Colorado courses averages $86, of which $9 is spent on automobile travel to the courses and $49 is paid for green and cart fees, leaving a consumer surplus of $28. Our analysis suggests that the total net economic value associated with the 7.8 million rounds of golf in Colorado is $224.64 million annually. Since this is the first study that estimates the consumer surplus for golfing we can only compare this surplus to other somewhat dissimilar recreation activities like hiking, which has a roughly equivalent consumer surplus. Given the popularity of golf and the fact that much of the activity occurs at municipal golf courses, which purposely attempt to keep fees lows, we suspect that there is an equivalent consumer surplus received by golfers throughout the nation.

References

Bishop, R. and T. Heberlein. 1979. Measuring Values of Extra-Market Goods: Are Indirect Measured Biased? American Journal of Agricultural Economics 61: 926-930.

Brown, W. and F. Nawas. 1973. Impact of Aggregation on the Estimation of Outdoor Recreation Demand Functions. American Journal of Agricultural Economics 55: 246-249.

Brown, W., C. Sorhus, B. Chou-Yang and J. Richards. 1983. Using Individual Observations to Estimate Recreation Demand Functions: A Caution. American Journal of Agricultural Economics 65: 154-157.

Cesario, F. and J. Knetsch (1970). Time Bias in Recreation Benefit Estimates. Water Resources Research 6(3): 700-704.

Clawson, M. and J. Knetsch. 1966. The Economics of Outdoor Recreation. Johns Hopkins University Press, Baltimore, MD.

Correia, A., C. Barros and A. Silvestre. (2007). Golf Tourism Repeat Behavior in the Algarve: A Mixed Logit Approach. Tourism Economics 13(1): 111-127.

Creel, M. and J. Loomis, 1990. Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California. *American Journal of Agricultural Economics*, Volume 72(2): 434-41.

Davies, S., P.Watson, A. Cramer, and D. Thilmany. 2004. The Economic Contribution to Colorado's Golf Industry. Economic Development Report, EDR-04-08, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO 80523.

Gum, R. and W. Martin. 1975. Problems and Solutions in Estimating the Demand for and Value Rural Outdoor Recreation. American Journal of Agricultural Economics 57: 558-566.

Hellerstein, D. (1992), "The treatment of non-participants in travel cost analysis and other demand models", Water Resources Research, Vol 28, No 8. pp. 1999-2004.

Imacination Software, (2002). "*Zip Distance Calculator*". Available at: http://webservices.imacination.com/distance/

Loomis, J. and R. Walsh. 1997. Recreation Economic Decisions, 2nd Edition. Venture Publishing Inc., State College, PA.

Milam, R. and E. Pasour, Jr. 1970. Estimating the Demand for an On-Farm Recreational Service. American Journal of Agricultural Economics 52(1): 127-131.

Mulligan, J. The Pricing of a Round of Golf: The Inefficiency of Membership Fees Revisited. Journal of Sports Economics 2(4): 328-40.

National Highway Institute (1995). "*Estimating the Impacts of Urban Transportation Alternatives, Participant's Notebook*", National Highway Institute, Federal Highway Adm. (Washington DC), Course # 15257, p VI-16

Shmanske, S. 2001. Price Discrimination at the Links. Contemporary Economic Policy 16: 368-378.

US Census Bureau (2000). "*Fact Sheet*". Available at: http://factfinder.census.gov/servlet/SAFFacts?_event

US Department of transportation, (2003) "Average Cost per Mile of Owning and Operating an Automobile: 1993-2003. Available at: answers@bts.gov

Wilson, J. (2005). "Golf and Agritourism: Exploring the Role of Natural Amenities in Western US Recreation Economies" Masters Thesis, Colorado State University, Fort Collins, Colorado.

SELECTION EFFECTS IN META-VALUATION FUNCTION TRANSFERS

Randall S. Rosenberger
Oregon State University

Robert J. Johnston
University of Connecticut

ABSTRACT

Seemingly independent influences on and choices in conducting and reporting primary research may emerge as biases in a stock of knowledge. Selection effects may arise from socio-political influences (*research priority selection*), researcher choices (*methodology selection*), peer-review influences (*publication selection*) and meta-analyst choices (*metadata sample selection*). We discuss these four types of selection effects including how to detect them, empirical evidence of them in the meta-analysis literature, and their implications for future benefit transfers and primary research. Meta-regression analysis may be our best tool for detecting and correcting these selection biases.

Acknowledgements

This research was supported in part by funds from the U.S. EPA STAR grant, "Meta-Regression Analysis of Recreation Valuation and Demand Elasticities: Identifying and Correcting Publication Selection Bias to Improve Benefit Transfer", #RD-832-421-01 to Oregon State University.

I. INTRODUCTION

Benefit transfer uses existing estimates of WTP derived from primary research to predict WTP for other sites of policy significance for which primary valuation estimates are unavailable. It may be described as the "practice of taking and adapting value estimates from past research … and using them … to assess the value of a similar, but separate, change in a different resource" (Smith, van Houtven and Pattanayak 2002, p. 134). Although the use of primary research studies to estimate values is almost universally preferred, the realities of the policy process often dictate that benefit transfer is the only option for assessing certain types of non-market values. Although meta-regression analysis may be used for a variety of analytical tasks in environmental economics[8], recent works have given increasing attention to the potential use of meta-analysis to inform function-based benefit transfer (Bergstrom and Taylor 2006; Johnston et al. 2005; Rosenberger and Stanley 2006). This attention is at least in part due to the increasing availability of empirical estimates of non-market value (i.e., willingness to pay or WTP) from which metadata may be constructed.

One of the primary advantages of meta-analysis as a benefit transfer tool relates to its capacity to allow more appropriate adjustments of welfare measures based on patterns observed in the literature. Within a benefit transfer context, transfer error is often inversely related to the correspondence between a study site and a policy site among various dimensions (Rosenberger and Phipps 2007). The probability of finding a good fit between a single (or multiple) study site and a policy site, however, is usually low (Boyle and Bergstrom 1992; Spash and Vatn 2006). If, on the other hand, empirical studies

[8] For example, meta analysis may be used to synthesize a body of literature or conduct hypothesis tests on the effects of moderator variables on measured effect sizes, among other uses.

contribute to a body of WTP estimates (i.e., metadata), and if empirical value estimates are systematically related to variations in resource, study and site characteristics, then meta-regression analysis may provide a viable tool for estimating a more universal transfer function with distinct advantages over unit value or other function-based transfer methods (Johnston, Besedin and Wardwell 2003; Rosenberger and Loomis 2003; Rosenberger and Stanley 2006). More specifically, Rosenberger and Phipps (2007) posit a meta-valuation function as the envelope of a set of empirically-defined valuation functions reported in the literature.

Despite the promise of such methods, the use of a meta-valuation function for benefit transfer assumes that the underlying body of valuation literature is a random, unbiased sample of the population of empirical estimates, and that these combined empirical estimates provide an unbiased representation of true, underlying resource values. Therefore, meta-valuation function transfers can only be as good as the data from which they are derived or to the extent that any measurable biases may be corrected prior to the transfer process.

This paper coordinates original empirical results with prior findings from the meta-analysis literature to elucidate issues, tradeoffs and concerns related to *selection effects* in meta-analysis benefit transfers. Selection effects of interest include those based on socio-political influences, researcher choices, peer-review influences, and meta-analyst choices. Although the nature of meta-analysis used as a benefit transfer tool together with the characteristics of the valuation literature imply that selection effects such as these may be pervasive, few meta-analyses test for, or address potential implications of such effects. Discussions of selection effects are sparse and scattered throughout the literature, leaving practitioners with no clear, consolidated guidance regarding ways in which such effects can manifest in meta-analysis, as well as possible ameliorative measures that might be taken. This paper endeavors to fill this gap in the literature, providing a more comprehensive discussion of the potential influences of selection effects on meta-analyses used for benefit transfer. This includes a review of the literature addressing such effects and a discussion of general insights that may be drawn from this literature.

We begin with conceptual discussions of primary issues, followed by illustrations of potential implications for non-market valuation and benefit transfer based on case-study metadata addressing values for a range of different natural resources. The discussion highlights related tradeoffs facing meta-analysts who seek to apply results for benefit transfer, the state-of-the-literature with regard to these tradeoffs, potential solutions to remaining concerns, and crucial areas for future research.

II. SELECTION EFFECTS, META-ANALYSIS, AND BENEFIT TRANSFER

Seemingly independent influences on and choices in conducting and reporting primary research may emerge as biases or systematic patterns in a stock of knowledge (metadata). Bias in a stock of knowledge may arise from numerous decisions made by a researcher, including those that determine (1) What issue to research (*research priority selection*); (2) How to research the issue (*methodology selection*); (3) What and how to report results (*publication selection*); and (4) How to meta-analyze the data (*sample selection*). If such patterns or biases are unanticipated by researchers, resulting selection effects can lead to unforeseen biases in benefit transfers. Meta-regression analysis provides a potential means to identify, measure, and correct for underlying biases in primary research—thereby improving the validity and accuracy of benefit transfer. In cases where needed corrections are either not made or are infeasible, however, biases in the underlying stock of knowledge (or literature) may carry over into empirical benefit transfers—including those conducted using meta-analysis.

Here, we emphasize four possible sources of selection effects, broadly defined. These are emphasized based on their likely pervasiveness in the valuation literature, and hence impacts on associated meta-analysis. *Research priority selection* may be driven by socio-political circumstances (societal awareness and perceived importance) of a particular resource, or the extent to which agencies are willing and able to fund research in a particular area. As a result, analyses may tend to target resources with higher marginal or total values, all else held constant (Hoehn 2006). *Methodology selection* affects WTP estimates and further complicates the use of meta-valuation function transfers when methodological characteristics are significant determinants of the variation in WTP (Johnston et al. 2005; Johnston, Besedin and Ranson 2006). Although covariates in statistical meta-regression models can quantify associated effects, the treatment of these effects for value prediction (e.g., in a benefit transfer context) may have significant implications for resulting estimates (Johnston, Besedin and Ranson 2006; Moeltner, Boyle and Paterson 2007).

Publication selection, either as a decision of the researcher or the peer-review process, is also known to bias a literature (Rosenberger and Stanley 2006; Stanley 2005). Due to journal, reviewer, and researcher publishing criteria, results of entire studies, standard datasets and estimated models, and specific information on study sites and sample populations may be suppressed. Even once a study is present in the literature, however, there is no guarantee that it will be selected by researchers implementing a meta-analysis. Potential selection effects related to choices made by meta-analysts in composing metadata are denoted *metadata sample selection*. Just as sample selection is known to bias estimates of value if not corrected in primary data models (Bateman et al., 2002; Garrod and Willis 1999), it is also a relevant concern for meta-regression models (Bergstrom and Taylor 2006; Moeltner and Rosenberger 2007). Defining the policy-relevant resource and identifying relevant studies (including as an attempt to avoid other selection effects) to be included in the metadata can affect the estimated meta-valuation function and predicted values that arise.

The application of meta-regression analysis to applied benefit transfer requires that researchers not only *identify patterns* associated with the above-noted selection effects, but also make *appropriate assumptions* regarding the treatment of these effects when applying statistical results to predict values. In some cases, theory provides clear guidance for these assumptions and treatments. In other cases, however, neither economic theory nor the literature provides significant guidance, leading to a situation in which ad hoc researcher decisions may have substantial impacts on benefit transfer results. In such cases, it is critical to identify the sensitivity of transfer estimates to researcher decisions.

III. A CONCEPTUAL MODEL FOR META-ANALYSIS BENEFIT TRANSFER

As a foundation for subsequent discussion, we begin with a simple conceptual model for function-based benefit transfer based on meta-regression results. Welfare measures for environmental resources are derived primarily by individuals' expressing their level of welfare based on tradeoffs observed through choices they either make (revealed preferences) or intend to make (stated preferences) (Champ, Boyle and Brown 2003; Freeman 2003; Garrod and Willis 1999). For each empirical study, an aggregate or mean measure of welfare for the representative individual in the study is reported. This aggregate welfare measure, $\bar{y}_{js}$, then becomes the measured effect size in a meta-regression model:

$$\bar{y}_{js} = \bar{x}_{js}\beta + \varepsilon_{js} . \qquad [1]$$

Thus, $\bar{y}_{js}$ is the welfare measure for site *s* in study *j*, $\bar{x}_{js}$ is a vector of variables measuring characteristics of site *s* of study *j*'s representative individual (age, income, experience, attitudes), environmental conditions for site *s* of study *j*, causes of environmental conditions for site *s* of study *j*, research methods used for site *s* in study *j*, temporal factors for study *j*, and locational factors for study *j*. We assume that ε_{js} is an i.i.d. distributed normal error term with zero mean, and the β's are the meta-regression coefficients to be estimated by the meta-regression model. We also assume there are no effects of study-specific unobservables and heteroskedasticity in the reported welfare measures.[9] Prediction of an aggregate welfare measure for the policy site, $\hat{y}_p$, using the meta-regression model simply replaces the moderator effects, $\bar{x}$'s, with measures at the policy site:

$$\hat{y}_p = x_p\hat{\beta} . \qquad [2]$$

[9] Although one could easily relax model assumptions to allow for such possibilities, which we do when discussing publication selection effects, doing so in the base model would not contribute significantly to the discussion presented here. Hence, we retain these simplifying assumptions for the sake of conciseness.

While a goal of meta-regression analysis is the isolation and measurement of moderator effects, which may of course be interpreted as conditional upon the existing metadata, applications of meta-regression analysis for benefit transfer must be cognizant of the broader context, history, and trends in primary research when it is gathered in a meta-dataset. Prior circumstances and decisions may significantly affect the quality, type and amount of empirical evidence available for an issue, as well as the interpretation of available evidence. Within a benefit transfer context, these issues influence both the potential for unbiased estimation of the β coefficients in [1], and also the appropriate choices for moderator effects x_p, in [2]. Unbiased benefit transfer depends on unbiased β coefficients and appropriate assignment of moderator effects—both of which are influenced by broader selection effects as noted above.

The following sections address the four primary areas of potential selection biases summarized above. In each case, we highlight potential impacts on the validity of function-based benefit transfer using meta-regression results. We also note corrective actions that may be taken in the short run to adjust for such selection effects, where such actions exist. Finally, we note broader research recommendations and needs that may ameliorate such problems in the longer run.

IV. RESEARCH PRIORITY SELECTION

If primary research is randomly distributed over resources, populations and policy contexts, then all occurrences have an equal probability of being sampled. Socio-political circumstances, however, can play a major role in whether a resource, population and/or policy context is studied. Those resources that are not evaluated have no observations on $\overline{y}_{js}$. If there is a research priority selection sampling bias, then we would expect high valued, more prominent occurrences of a resource to be evaluated, while more mundane (perhaps lower value) occurrences of a resource would be overlooked. Hoehn (2006) identifies four plausible descriptors that are correlated with the decision to conduct primary research on a resource and the availability of estimated effect sizes for the resource. The probability of conducting primary research increases with increases in our awareness of the resource, the importance of the resource to stakeholders, the magnitude of the policy decisions to be made in response to conflicts over the resource, and the availability of funding to support primary research.

In this case, relatively standard methods may be used to adjust meta-analysis results. To adjust for resulting research priority selection bias of missing effect sizes, Hoehn (2006) proposed a two-stage Heckman sample selection model. This model has also been proposed for evaluating publication selection bias (Florax 2002; Smith and Huang 1993), which we will deal with in a later section of this paper. To implement the corrective model we define a latent or unobserved variable, z^*_{js}, that defines a threshold that is crossed when $\overline{y}_{js}$ is reported. This latent variable may be defined by an identifiable process (Greene 2003):

$$z^*_{js} = \overline{w}_{js}\alpha + \mu_{js} \qquad [3]$$

where $\overline{w}_{js}$ is a vector of variables explaining the selection process for site s of study j, α is a vector of coefficients to be estimated in the selection equation, and μ_{js} is an i.i.d. distributed normal error term with zero mean. In the Heckman bivariate model that links equations [1] and [3], the error terms are distributed as bivariate normal:

$$\varepsilon, \mu \sim N\left(0,0,\sigma_{\varepsilon}^{2},\sigma_{\mu}^{2},\rho\right) \qquad [4]$$

with zero means and correlation ρ.

By definition, our latent variable z^{*} is not observed, but its counterpart z is observed when a threshold is crossed; that is:

$z = 1$ if $z^{*} > 0$; and

$z = 0$ if $z^{*} \leq 0$. [5]

Therefore, $\bar{y}$ and $\bar{x}$ are observed when $z = 1$. Given the non-zero covariance between ε and μ, consistent estimates of moderator effects, β, are given by:

$$E[\,\bar{y}_{js} \mid \bar{y}_{js} \text{ is observed}] = \bar{x}_{js}\beta + \beta_{\lambda}\lambda_{js}\left(\alpha_{\mu}\right), \qquad [6]$$

where $\alpha_{\mu} = -\,\bar{w}_{js}\alpha/\sigma_{\mu}$ and the inverse Mills ratio $\lambda\left(\alpha_{\mu}\right) = \phi\left(\bar{w}_{js}\alpha \big/ \sigma_{\mu}\right) \Big/ \Phi\left(\bar{w}_{js}\alpha \big/ \sigma_{\mu}\right)$.

In the case where no selection effects are measured, β_{λ} is zero and β's in equation [1] are consistent and unbiased. In the case where selection effects are measured, then λ_{js} becomes our omitted variable in equation [1]. Extensions to panel data models that deal with intra-study correlations when metadata includes studies that report more than one welfare estimate per resource investigated are relatively straightforward (e.g., Greene 2002; Hoehn 2006).

Hoehn's (2006) application of the above model seeks to quantify research priority selection in the wetland valuation literature. If observations on wetland values were based on a random sample, then they should have equal probabilities of being selected and β_{λ} in equation [7] would be zero. Hoehn (2006) hypothesized there may be a jurisdictional bias in how wetlands are selected for research. Therefore, he conducted a random effects Heckman model at the state level for the US. The explanatory variables in the selection equation [3] were defined to include the ratio of wetland acres to total open space acres, population density and per capita income for each state. The ratio of wetland acres to open space acres was the only statistically significant variable in the selection equation. It shows that larger ratios are likely associated with increased awareness of the resource and that development is more likely to affect wetlands. The inverse Mills ratio was positive and statistically different than zero, indicating selection bias in this body of literature. The coefficient estimates in the Heckman corrected model for the most part decrease in absolute magnitude from the uncorrected OLS model. The uncorrected generic wetland value is four times larger than the Heckman estimate, showing a substantial research priority selection bias.

In the absence of a model that explicitly corrects for research priority selection, potential *preliminary* indicators of selection bias may include the correlation between observations on resource values and time.[10] The rationale for a time-trend indicator is that if high valued resources are selected for primary research first, we would expect a significant and negative association with resource values over time. Table 1 shows a summary of several studies that tested and reported results for a trend parameter in their specified valuation meta-analyses. In some meta-analyses, this trend effect ($\beta_T < 0$) has been found to be significant and negative (Johnston, Besedin and Wardwell 2003; Johnston et al. 2005 for surface water quality and aquatic habitat resources, respectively), as hypothesized.[11] Smith and Kaoru (1990b) also show travel cost own-price elasticity of demand estimates were becoming more elastic over time, possibly signaling a trend

[10] We thank John Loomis for this argument.
[11] An alternative explanation for the negative trend in values over time for a resource is advances in methods that minimize bias (Johnston et al. 2005), although this assumes biases nearly always lead to larger value estimates.

from high quality, unique sites or single site models to lower quality, substitutable or regional models, *inter alia*, all of which would lead to more elastic own-price elasticity of demand estimates (Loomis and Walsh 1997).

In the recreation valuation literature, however, the opposite is often found on trend variables—the coefficient on a trend variable ($\beta_T > 0$) is statistically significant and positive (Bateman and Jones 2003; Johnston et al. 2006; Rosenberger and Loomis 2000a; Rosenberger and Stanley 2007). This may indicate recreation values are growing over time at a rate greater than selection biases that manifest as temporal trends. Or, the importance of recreation resources was extensive enough to warrant estimation of general, baseline recreation values regardless of site quality or socio-political influences, as is the case for nationally-scoped, agency-driven research conducted by the US Forest Service and US Fish and Wildlife Service. Therefore, the recreation values literature as a whole may not suffer from research priority selection effects, but for particular activities, such as mountain biking, it may. In yet other meta-analyses, trend effects ($\beta_T = 0$) were not found to be statistically significant (Loomis and White 1996; Woodward and Wui 2001, for endangered species and international wetland resources, respectively).

In sum, evidence strongly suggests that research priority selection exists in at least some areas of the non-market valuation literature—implying that models that account for sample selection may be necessary to prevent associated biases in benefit transfers. Evidence also suggests, however, that the prevalence of such patterns may vary across different types of resources and value-generating activities.

V. METHODOLOGY SELECTION

Researchers conducting primary studies may choose from among several accepted models and methods when estimating economic values of resources (Champ, Boyle and Brown 2003; Freeman 2003; Garrod and Willis 1999). Theory typically indicates that WTP should *not* vary according to methodological attributes—with the exception of those that would cause different components or types of WTP to be estimated (e.g., use WTP only versus a combination of use and nonuse WTP; stated WTP estimated under different information sets) (Johnston, Besedin and Wardwell 2003). For example, while theory suggests that WTP values for otherwise identical resource improvements should *not* be convergent for most revealed and stated preference analyses—because these two methodological categories generally estimate theoretically distinct welfare measures—otherwise identical stated preference welfare effects should in theory be convergent, irrespective of such features as the type of stated preference survey implementation applied (e.g., mail versus in-person surveys).

Notwithstanding such theoretical expectations, methodological choices often result in systematic effects on estimated values. These systematic effects become visible in meta-regression models, which have found methodology attributes of primary studies to influence estimated WTP values, including study type, survey implementation method, response rate, question format, treatment of outliers/protests, and econometric methods, *inter alia* (Bateman and Jones 2003; Brouwer et al. 1999; Johnston, Besedin and Wardwell 2003; Johnston et al. 2005, 2006; Rosenberger and Loomis 2000a,b; Poe, Boyle and Bergstrom 2001; Smith and Osborne 1996).

The estimation of these systematic effects has been one of the goals of recent meta-regression analysis in the environmental economics literature (Johnston, Besedin and Ranson 2006). The concern, however, is how to treat these methodological effects when using the meta-regression model to predict estimates for a policy site. We expand our meta-regression model of equation [1] to account for methodology attributes:

$$\bar{y}_{js} = \bar{x}_{js}\beta + m_{js}\beta_m + \varepsilon_{js} \qquad [7]$$

where m_{js} is a vector of methodology attributes (typically modeled as dummy variables indicating the use of particular methods) and β_m are associated coefficients to be estimated. Estimation of equations such as [7] is straightforward. The difficulty for applied benefit transfer is determining *what value* to insert for m_p in the empirical transfer function:

$$\hat{y}_p = x_p\hat{\beta} + m_p\hat{\beta}_p\,. \qquad [8]$$

Past benefit transfers have typically ignored these effects (which risks omitted variable bias); used ad hoc adjustments such as the mean level ($\overline{m}$) from the metadata, which holds the effect constant at its mean level (Rosenberger and Loomis 2000a, 2003); or otherwise suppressed information regarding the sensitivity of WTP to methodology attributes. Johnston, Besedin and Ranson (2006), however, show that WTP predictions in a meta-regression function transfer can be highly sensitive to the analyst's treatment of methodological attributes—confidence intervals can vary by a factor of fifteen when ad hoc treatments of methodology attributes are drawn from single studies.

Ad hoc treatments may also be guided by theory, expectations, or analysts' beliefs regarding the reliability and validity of individual valuation methods. For example, revealed preference methods such as the travel cost method are less controversial to the broader economic profession simply because they are based on real choices that are observable. However, within revealed preference methods, researchers' choices regarding the treatment of the value of time, out-of-pocket costs of travel, among other issues, have systematic effects on values and remain controversial. Other choices are driven by theory such as including the price of substitute sites in travel cost models. However, in most cases these treatments require the analyst to accept relatively strong assumptions that may not be shared within the research or policy community.

Another approach to methodological sensitivity is to 'hold constant' the effect of methodology *in the metadata sample* by setting the methodological attributes in the transfer application at their mean values for the dataset. If the metadata represent a random sample of the empirical studies, then using the mean value of the metadata may be an unbiased representation of the use of different methods in the literature. This is especially true for metadata sets with a large number of observations, following large sample theory properties (Johnston, Besedin and Ranson 2006).

Recent work also suggests that the degree of transfer error resulting from mean value treatments of methodological covariates may be relatively modest. Stapler and Johnston (2007) implement an out-of-sample validity test that characterizes the systematic impact of methodological covariate treatment on transfer error. Using repeated leave-one-out (jackknife) cross-validation, the analysis contrasts errors for a hypothetical ideal case in which correct methodological covariate treatments are known (i.e., methodological variable levels are assigned to match those of each out-of-sample test case) to the less-than-ideal but realistic case in which the correct treatment of these covariates is unknown and mean values are used. Results suggest that the additional error associated with the common mean value treatment of methodological covariates is relatively modest, on average.

Such findings notwithstanding, the use of mean values for methodological attributes may also ignore temporal trends in methodological developments and applications. Over time, new or improved methods may be introduced that are subsequently tested and/or adopted by the valuation research community, leading to a trend in the application and use of primary methods. For example, in the recreation valuation literature, early applications of revealed preference methods primarily used a zonal travel cost method, but then have switched to individual travel cost methods with an increase in the application of random utility models. In stated preference applications, the early favorite was open-ended elicitation formats, subsequently giving way to dichotomous choice applications (Rosenberger and Stanley 2007). These temporal trends have significant effects on the magnitude of values being estimated in the literature.

Just as primary researchers must make choices regarding methods to be used, transfer analysts must also make choices on how to treat systematic methodology effects in meta-regression models. Each approach has pros and cons, and thus far, the literature provides relatively minimal guidance on the issue of how to treat the often significant methodology effects in a benefit transfer setting. This represents a significant area for future research with critical implications for applied benefit transfer.

VI. PUBLICATION SELECTION

Publication selection bias arises when a literature is not an unbiased sample of empirical evidence. Card and Kreuger (1995, p.239) identify three potential sources of publication selection in economics—(1) reviewers and editors may be predisposed to accept papers consistent with the conventional view; (2) researchers may use the presence of conventionally expected results as a model selection test; and (3) everyone may possess a predisposition to treat 'statistically significant' results more favorably. Smith and Pattanayak (2002, p.273) identify another plausible source of publication selection bias—(4) most journals in the environmental economics field are not interested in new estimates of benefits for their own sake. These journals may be predisposed to select manuscripts based primarily upon methodological innovations and contributions. When a literature is evaluated for publication selection bias, it is often found and may have substantial impacts on inferences derived from the literature (Rosenberger and Stanley 2006). For example, price elasticities of water demand have been found to be exaggerated four-fold through publication selection bias (Dalhuisen et al. 2003; Stanley 2005). The primary concern for benefit transfer is that a publication selection process may suppress information that, while not considered sufficiently important for publication, may be highly relevant for benefit transfer.

A Heckman two-stage sample selection model as presented earlier has been suggested as a potential model for analyzing publication selection (Florax 2002; Smith and Huang 1993; Stanley 2006), along with several other parametric and nonparametric methods (Florax 2002; Stanley 2005, 2006). However, an additional complication in the case of publication selection is that the $z = 0$ in equation [5] are likely not observed. That is, studies that go unreported are ultimately unknown—none of the $\overline{w}_{js}$ are observed for unreported studies. This means the inverse Mills ratio cannot be estimated and selection effects cannot be measured and corrected in the Heckman two-stage model as outlined above. Stanley (2006), however, proposes a model that relies on the inherent heteroskedasticity of effect sizes as a means to provide a proxy for the inverse Mills ratio. Studies with larger sample sizes and thus smaller standard errors of effect sizes will have a higher probability of being reported.

Another approach to publication bias considers the various channels through which information may be provided—for example in the peer reviewed versus grey literature. Such approaches cannot address patterns related to research results that go unreported in any format, but can address patterns of selection bias in, for example, outlets such as academic journals. One of the primary mechanisms used to detect publication selection bias of the latter type has been through the use of dummy variables in a meta-regression model (Table 2). This approach typically includes an additional moderator variable, q_{js}, which identifies how the results of a study are made available:

$$\bar{y}_{js} = \bar{x}_{js}\beta + m_{js}\beta_m + q_{js}\beta_q + \varepsilon_{js}, \qquad [9]$$

where β_q are the added coefficients to be estimated. One of the attributes of this meta-regression model is its robustness to observable primary study misspecifications and methodological choices.

Based on such an approach, Koetse, Florax and de Groot (2005, p. 1) argue that "the current practice of accounting for such primary study aberrations in a meta-analysis by means of dummy variables goes a long way in mitigating their negative effects on the bias and mean squared error of the estimator, and the size and the power of the statistical tests on the meta-estimate." Publication selection moderator variables may be defined across the types of documents in the literature (e.g., journal article, agency report, consulting report, MS thesis or PhD dissertation, proceedings paper, working paper, etc.), the peer-review process, or motivations of documents (e.g., introduce new estimator, improve efficiency, reduce bias, or introduce new estimates of value).

A similar approach is illustrated by Smith and Huang (1993), who apply a Heckman two-stage model for publication selection bias. They hypothesize that conventionally-expected results would tend to be published in the peer-reviewed literature (primarily journal articles), while other studies will not be published. They do not find statistically significant correlations between published and unpublished studies in their model, which may have been due to limited access to all empirical studies available, in particular when these studies are not published in the mainstream literature.

Table 2, in contrast, shows several meta-analyses that use a dummy variable approach in their meta-regression model. In most cases, publication dummy variables were statistically significant. For example, van Kooten et al. (2004) found peer-reviewed studies reported higher carbon sequestration cost estimates than non-peer-reviewed studies. Such results may be interpreted in opposite ways—the peer-review process is working by validating the design, analysis and interpretation, and ensuring the comparability across manuscripts surviving this process, or, due to selection effects as noted above, valid studies that use standard, accepted methods are not entering the peer-reviewed, published literature domain. Furthermore, different classes of journals may have different motives (introduction of new estimators versus the introduction of new estimates of values) that lead to different publication selection criteria.

For example, Gallett and List (2003) (elasticities of cigarette demand) included a dummy variable identifying publication in the top 36 economics journals. This measure of journal prestige was significant and negative in the price elasticity model and significant and positive in the income elasticity model. Both of these directional effects imply that reported demand elasticities are larger (more elastic) in the most prestigious economics journals than other outlets for publishing environmental research results. The policy implications are significant—in one publishing venue cigarette demand would be relatively more responsive to policy, while in another venue it would be significantly less responsive. This is precisely the sort of effects that are the expected result of publication bias. Publication pressures cause researchers and/or reviewers to use theoretical or other prior expectations to select among submitted results. The higher the prestige of the journal, the greater this selection bias may be.

Rosenberger and Stanley (2007) provide preliminary tests of publication selection bias in the recreation use values literature. They use t-tests and meta-regression analysis to identify publication effects based on document type and publication motivations. The analysis finds that journals, agency/university reports, PhD dissertations, working papers and proceedings papers do not provide statistically different mean values of WTP from each other, but do provide mean values statistically different from books/book chapters, consulting reports, and Master's theses. Mean values provided by consulting reports differed significantly (in a positive direction) from all other document types. All document types with the exception of books/book chapters are statistically different than zero, meaning they are statistically different from the omitted category of Master's thesis. Primary contribution effects only show that efficiency contributions differ from zero and from the omitted category of new values. Bias testing and new values are not distinguishable from each other.

In sum, although the current literature does not suggest a single “fix” for publication selection bias, various avenues are available that provide at least potential amelioration for suspected effects. These include Heckman-type selection biases where feasible, as well as simpler approaches that adjust for systematic patterns in WTP associated with various publication avenues (e.g., using dummy variables in meta-regression). All these methods, however, involve comparisons among various different venues of publication (e.g., peer reviewed, non-peer reviewed). A continuing challenge is the identification of biases in results published in *any venue*, relative to results that go entirely unreported. Although a small number of preliminary approaches have been suggested (e.g., Stanley 2006, 2007), this remains an important area for future work.

VII. METADATA SAMPLE SELECTION

Thus far, the above-noted effects and biases have all related to patterns of research results found in the broader literature. The last source of potential bias, in contrast, relates to choices of the meta-analyst regarding which studies to incorporate in a particular set of metadata.

One of the most obvious instances in which meta-analyst sample decisions are critical is in choosing which particular resources or policy contexts are appropriate for inclusion within the metadata. In a benefit transfer setting, one of the first steps is to define the policy context for which transfer values are needed (Bergstrom and Taylor 2006; Boyle and Bergstrom 1992; Rosenberger and Loomis 2003; Stanley 2001). Defining the policy context entails an implicit determination of those studies considered relevant; i.e., the optimal scope of the metadata. This question of optimal scope is important when conducting meta-regression function transfers. In econometric terms, the question of optimal scope can be interpreted as the exact definition of the dependent variable in the meta-regression model, which, in turn, defines the set of source studies to be considered for inclusion in the metadata. This issue has been briefly raised at various points in time in the literature (e.g., Bergstrom and Taylor 2006; Segerson 1994; Engel 2002), but has not been examined in depth in existing contributions. The primary tradeoff is often between maintaining close similarity among dependent variables versus including additional information (i.e., observations) in the metadata.

Moeltner and Rosenberger (2007) used Bayesian model search and model averaging techniques to illustrate how these methods may better utilize existing information on values for benefit transfer predictions and help define the optimal scope of a metadata set. Within this study, they define the policy context as consisting of a need for an estimate of average welfare per day for access to a site for coldwater fishing in a running water environment. Their metadata, under these restrictive selection conditions, comprised 15 studies providing a combined total of 73 estimates of value (further restricted to single-site models in the US). Possible scope augmentation of this limited meta-dataset includes warm water fisheries and still water environments. If the metadata is augmented on these two dimensions, then the dataset increases to 37 studies with a combined 229 observations. They found that the meta-regression function increased in efficiency when the baseline data were augmented along the dimension of warm water fishing, but not along the dimension of still water environment. Therefore, in defining the policy context, the transfer model was improved by expanding the scope of relevant studies to include other types of fish species (coldwater and warm water), while holding the type of environment (running water) constant. This illustrates that meta-analysts' decisions regarding the appropriate scope of metadata may lead to sample selection bias in the resulting meta-regression model and subsequent benefit predictions for the policy site.

Beyond sampling issues related to the specific resources and policy contexts addressed by original research studies, meta-analysts may also decide to attenuate certain selection biases by restricting the scope of studies to be included in the metadata. But this may only exacerbate the above-noted selection biases through additional sample selection issues. For example, one may conclude that results reported in top-tier journals, because they survive a strenuous selection process, provide the most reliable, rigorous estimates of values. However, the selection criteria of these journals may not be complementary to criteria for selecting studies that provide good estimates of value for benefit transfer (see publication selection discussion above). In another case, metadata might be restricted to a particular geographic region citing environmental and social consistency within a region. But, as Rosenberger and Loomis (2000a) show, a meta-analysis transfer function derived from combining data from all regions outperformed regional models in in-sample convergent validity testing.

Similarly, one might narrow the metadata to valuation studies that use a particular valuation approach (e.g., stated preference methods, as in Johnston et al., 2005; or travel cost model estimates, as in Smith and Kaoru 1990a). While such approaches may improve the statistical fit of estimated models and may be justified for theoretical or other reasons (Smith and Pattanayak 2002), they may also magnify any biases that may be present in the underlying literature. As above, this is an area in which the literature provides little clear guidance, and in which future research is needed to address implications for benefit transfer and potential tests or appropriate corrective measures that might be undertaken.

VIII. CONCLUSIONS

Benefit transfer is an almost universal component of most large-scale benefit cost analyses, and reliance on benefit transfer to estimate resource values for policy purposes is only expected to increase. Iovanna and Griffiths (2006), for example, expect the US EPA, which has applied benefit transfers in many prior cases, to increase their use of benefit transfer due to its expediency, agency-financial constraints, and administrative hurdles associated with primary research. Meta-regression models have substantial potential for use in benefit transfers (Johnston et al. 2005; Rosenberger and Loomis 2003), but not without their own set of challenges.

The benefit transfer literature has provided promising findings with regard to the ability of meta-analysis to provide appropriate mechanisms for benefit transfer in some cases (Bergstrom and Taylor 2006). Evidence presented here also suggests, however, that researchers should consider the potential for selection effects when conducting applied benefit transfer of any type—including that using meta-analysis. Any research literature is characterized by numerous latent factors or patterns. Many of these only emerge when looked at through a meta-analysis lens. Benefit transfer analysts should be cognizance of these latent factors and in many cases the identification of methods to adjust for resulting patterns in WTP. Research results reviewed above show that the output of benefit transfer can be highly sensitive to various manifestations of selection effects—with significant implications for the validity of associated benefit transfers.

While this paper has emphasized the role of selection effects in meta-analysis, we stress that identical effects can apply to any application of benefit transfer. The primary difference between meta-analytic and non-meta-analytic methods in this regard is that meta-analysis can in many cases render resulting patterns and biases explicit (e.g., can quantify their magnitude and implications for benefit transfer). In contrast, alternative methods for benefit transfer leave selection effects largely unquantified and uncorrected, even when such effects are present. Hence, meta-analysis offers—at least in some cases—an ability to correct for selection effects that is not present in other mechanisms for benefit transfer.

The appropriate treatment of selection effects—and hence the ability of meta-analysts to take corrective measures—of course varies by the type of selection effect considered. Some of these effects, once known, may be offset or adjusted using standard statistical methods. Others may only be attenuated by new directions in primary research, such as estimation of baseline values for resources. Still other effects, however, are the direct result of judgments made by meta-analysts and benefit transfer practitioners. Sensitivity analyses of judgments and model assumptions should be provided to gauge their overall effects on benefit transfer accuracies.

This paper does not claim to identify all potential selection effects and possible solutions; instead its goal has been to provide a summary of effects likely to be most problematic and pervasive for applied benefit transfer. We hope that by raising researchers' awareness of potential selection effects, they may either avail themselves of possible solutions or at least note the possibility of WTP biases in resulting transfers. Within this context, it is important to note that the proposed "corrective" methods for selection effects—where possible—do not solve the problem of underlying biases within the research literature, nor are they a panacea for the many challenges in the use of welfare or WTP functions for benefits transfer. However, it is hoped that increased use of, and familiarity with such approaches can render more transparent the potential implications of selection effects for WTP, and thereby promote more informed and defensible applications of benefits transfer to environmental policymaking.

Finally, we emphasize that while results from selection effects testing in meta-analysis may provide some guidance to primary research in what and how it is conducted, the larger issue relates to the broader availability of quality primary research. Ideally, greater resources would be allocated to valuation research, encouraging a broader and more representative research literature. One might also encourage journals and other outlets to publish empirical valuation results that may not—due to a lack of methodological novelty or interesting "twists"—be easily accepted within typical academic outlets. These and other long term solutions, however, are costly (Loomis and Rosenberger 2006). With expected declines in research budgets and increased hurdles to primary research, we do not expect metadata to expand very quickly or in a directed fashion. In the absence of such broad changes in research approaches, it is critical for researchers to be aware of the context from which existing metadata are derived when conducting and applying meta-analysis for benefit transfer.

References

Bateman, I.J. and A.P. Jones. 2003. "Contrasting Conventional With Multi-Level Modeling Approaches to Meta-Analysis: Expectation Consistency in U.K. Woodland Recreation Values." *Land Economics* 79(2):235-258.

Bateman, I.J., R.T. Carson, B. Day, M. Hanemann, N. Hanley, et al. 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham, UK: Edward Elgar.

Bergstrom, J.C. and L.O. Taylor. 2006. "Using Meta-Analysis for Benefits Transfer: Theory and Practice." *Ecological Economics* 60(2):351-360.

Boyle, K.J. and J.C. Bergstrom. 1992. "Benefit Transfer Studies: Myths, Pragmatism and Idealism." *Water Resources Research* 28(3):657-663.

Card, D. and A.B. Kreuger. 1995. "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review* 85:238-243.

Champ, P.A., K.J. Boyle and T.C. Brown (eds.). 2003. *A Primer on Nonmarket Valuation*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dalhuisen, J.M., R.J.G.M. Florax, H.L.F. de Groot and P. Nijkamp. 2003. "Price and Income Elasticities of Residential Water Demand: A Meta-Analysis." *Land Economics* 79(2):292-308.

Engel, S. 2002. "Benefit Function Transfer Versus Meta-Analysis as Policy Tool: A Comparison." In R.J.G.M. Florax, P. Nijkamp and K.G. Willis (eds.), *Comparative Environmental Economic Assessment*. Cheltenham, UK: Edward Elgar.

Florax, R.J.G.M. 2002. "Methodological Pitfalls in Meta-Analysis: Publication Bias." In R.J.G.M. Florax, P. Nijkamp and K.G. Willis (eds.), *Comparative Environmental Economic Assessment*. Cheltenham, UK: Edward Elgar.

Freeman, A.M. 2003. *The Measurement of Environmental and Resource Values*, Second Edition. Washington, DC: Resources for the Future Press.

Gallett, C.A. and J.A. List. 2003. "Cigarette Demand: A Meta-Analysis of Elasticities." *Health Economics* 12(10):821-835.

Garrod, G. and K.G. Willis. 1999. *Economic Valuation of the Environment: Methods and Case Studies*. Cheltenham, UK: Edward Elgar.

Greene, W.H. 2002. *LIMDEP Version 8.0, Econometric Modeling Guide, Volume 2*. Plainview, NY: Econometric Software, Inc.

Greene, W.H. 2003. *Econometric Analysis*. Fifth Edition. Upper Saddle River, NJ: Prentice Hall.

Hoehn, J.P. 2006. "Methods to Address Selection Effects in the Meta Regression and Transfer of Ecosystem Values." *Ecological Economics* 60(2):389-398.

Iovanna, R. and C. Griffiths. 2006. "Clean Water, Ecological Benefits, and Benefits Transfer: A Work in Progress at the U.S. EPA." *Ecological Economics* 60(2):473-482.

Johnston, R.J., E.Y. Besedin and M.H. Ranson. 2006. "Characterizing the Effects of Valuation Methodology in Function-Based Benefits Transfer." *Ecological Economics* 60(2):407-419.

Johnston, R.J., E.Y. Besedin and R.F. Wardwell. 2003. "Modeling Relationships between Use and Nonuse Values for Surface Water Quality: A Meta-Analysis." *Water Resources Research* 39(12):1363-1372.

Johnston, R.J., E.Y. Besedin, R. Iovanna, C.J. Miller, R.F. Wardwell and M.H. Ranson. 2005. "Systematic Variation in Willingness to Pay for Aquatic Resource Improvements and Implications for Benefit Transfer: A Meta-Analysis." *Canadian Journal of Agricultural Economics* 53(2-3):221-248.

Johnston, R.J., M.H. Ranson, E.Y. Besedin and E.C. Helm. 2006. "What Determines Willingness To Pay Per Fish? A Meta-Analysis of Recreational Fishing Values." *Marine Resource Economics* 21:1-32.

Koetse, M.J., R.J.G.M. Florax and H.L.F. de Groot. 2005. *Correcting for primary study misspecifications in meta-analysis*. Tinbergen Institute Discussion Paper #TI 2005-029/3. Amsterdam, The Netherlands: Tinbergen Institute.

Loomis, J.B. and D.S. White. 1996. "Economic Benefits of Rare and Endangered Species: Summary and Meta-Analysis." *Ecological Economics* 18(3):197-206.

Loomis, J.B. and R.G. Walsh. 1997. *Recreation Economic Decisions: Comparing Benefits and Costs*, Second Edition. State College, PA: Venture Publishing, Inc.

Loomis, J.B. and R.S. Rosenberger. 2006. "Reducing Barriers in Future Benefit Transfers: Needed Improvements in Primary Study Design and Reporting." *Ecological Economics* 60(2):343-350.

Moeltner, K., K.J. Boyle and R.W. Paterson. 2007. "Meta-Analysis and Benefit Transfer for Resource Valuation—Addressing Classical Challenges with Bayesian

Modeling." *Journal of Environmental Economics and Management* 53(2):250-269.

Moeltner, K. and R.S. Rosenberger. 2007. *Meta-Regression and Benefit Transfer: Data Space, Model Space, and the Quest for 'Optimal Scope'*. Working paper. Reno, NV: University of Nevada, Reno, Department of Resource Economics.

Poe, G.L., K.J. Boyle and J.C. Bergstrom. 2001. "A Preliminary Meta Analysis of Contingent Values for Ground Water Quality Revisited." In J.C. Bergstrom, K.J. Boyle and G.L. Poe (eds.), *The Economic Value of Water Quality*. Northampton, MA: Edward Elgar Publishers.

Rosenberger, R.S. and J.B. Loomis. 2000a. "Using Meta-Analysis for Benefit Transfer: In-Sample Convergent Validity Tests of an Outdoor Recreation Database." *Water Resources Research* 36(4):1097-1107.

Rosenberger, R.S. and J.B. Loomis. 2000b. "Panel Stratification in Meta-Analysis of Economic Studies: An Investigation of Its Effects in the Recreation Valuation Literature." *Journal of Agricultural and Applied Economics* 32(3):459-470.

Rosenberger, R.S. and J.B. Loomis. 2001. *Benefits Transfer of Outdoor Recreation Use Values*. Gen. Tech. Report RMRS-GTR-72. Fort Collins, CO: USDA Forest Service, Rocky Mountain Research Station.

Rosenberger, R.S. and J.B. Loomis. 2003. "Benefit Transfer." In. P.A. Champ, K.J. Boyle and T.C. Brown (eds.), *A Primer on Nonmarket Valuation*, Dordrecht, The Netherlands: Kluwer Academic Publishers. Pp.445-482.

Rosenberger, R.S. and T.D. Stanley. 2006. "Measurement, Generalization, and Publication: Sources of Error in Benefit Transfers and Their Management." *Ecological Economics* 60(2):372-378.

Rosenberger, R.S. and T.D. Stanley. 2007. *Publication Effects in the Recreation Use Value Literature: A Preliminary Investigation*. Working Paper. Corvallis, OR: Oregon State University, Department of Forest Resources.

Rosenberger, R.S. and T.T. Phipps. 2007. "Correspondence and Convergence in Benefit Transfer Accuracy: Meta-Analytic Review of the Literature." In S. Navrud and R. Ready (eds.), *Environmental Value Transfer: Issues and Methods*. Dordrecht, The Netherlands: Springer. Pp.23-43.

Segerson, K. 1994. "The Benefits of Groundwater Protection: Discussion." *American Journal of Agricultural Economics* 76:1076-1078.

Smith, V.K. and J. Huang. 1993. "Hedonic Models and Air Pollution: Twenty-Five Years and Counting." *Environmental and Resource Economics* 3(4):381-394.

Smith, V.K. and J. Huang. 1995. "Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models." *Journal of Political Economy* 103(1):209-227.

Smith, V.K. and L. Osborne. 1996. "Do Contingent Valuation Estimates Pass the Scope Test? A Meta-Analysis." *Journal of Environmental Economics and Management* 31(3):287-301.

Smith, V.K. and S.K. Pattanayak. 2002. "Is Meta-Analysis a Noah's Ark for Non-Market Valuation?" *Environmental and Resource Economics* 22:271-296.

Smith, V.K. and Y. Kaoru. 1990a. "Signals or Noise? Explaining the Variation in Recreation Benefit Estimates." *American Journal of Agricultural Economics* 42(2):419-433.

Smith, V.K. and Y. Kaoru. 1990b. "What Have We Learned since Hotelling's Letter? A Meta-Analysis." *Economics Letters* 32:267-272.

Smith, V.K., G.van Houtven and S.K. Pattanayak. 2002. "Benefit Transfer Via Preference Calibration: "Prudential Algebra" for Policy." *Land Economics* 78(1):132-152.

Spash, C.L. and A. Vatn. 2006. "Transferring Environmental Value Estimates: Issues and Alternatives." *Ecological Economics* 60(2):379-388.

Stanley, T.D. 2001. "Wheat from Chaff: Meta-Analysis as Quantitative Literature Review." *Journal of Economic Perspectives* 15(3):131-150.

Stanley, T.D. 2005. "Beyond Publication Bias." *Journal of Economic Surveys* 19(3):309-345.

Stanley, T.D. 2006. *Two-Stage Precision-Effect Estimation and Heckman Meta-Regression for Publication Selection Bias*. Working Paper 1006/25. Victoria, Australia: Deakin University, School of Accounting, Economics and Finance.

Stanley, T.D. 2007. "Meta-Regression Methods for Detecting and Estimating Empirical Effect in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* (forthcoming).

Stapler, R.W. and R.J. Johnston. 2007. *Meta-Analysis, Benefit Transfer, and Methodological Covariates: Implications for Transfer Error*. Working Paper, Department of Agricultural and Resource Economics, University of Connecticut.

Van Kooten, G.C., A.J. Eagle, J. Manley and T. Smolak. 2004. "How Costly are Carbon Offsets? A Meta-Analysis of Carbon Forest Sinks." *Environmental Science Policy* 7:239-251.

Woodward, R.T. and Y. Wui. 2001. "The Economic Value of Wetland Services: A Meta-Analysis." *Ecological Economics* 37(2):257-270.

Zelmer, J. 2003. "Linear Public Goods Experiments: A Meta-Analysis." *Experimental Economics* 6(3):299-310.

TABLE 1
TREND EFFECTS IN VALUATION META-ANALYSES

Source	Resource	Trend Coefficient	Annual Effect
		$\beta_T < 0$	
Johnston et al. (2003)	Surface Water Quality Values	-0.0562 to -0.0735[a]	-\$1.06 to -\$1.08
Johnston et al. (2005)	Aquatic Habitat Values	-0.1058 to -0.1220[b]	-\$1.11 to -\$1.13
Smith & Kaoru (1990b)	Recreation Demand Price Elasticities	-0.42 to -0.52[c]	---
		$\beta_T > 0$	
Rosenberger & Loomis (2000a)	Recreation Values	1.161 to 1.246[d]	\$1.16 to \$1.25
Rosenberger & Stanley (2007)	Recreation Values	0.0138	\$1.01
Johnston et al. (2006)	Marginal Value per Fish	0.0875 to 0.1752[e]	\$1.09 to \$1.19
Bateman & Jones	Forest Recreation Values in UK	0.071 to 0.0755[f]	£0.08
		$\beta_T = 0$	
Woodward & Wui (2001)	International Wetland Values	-0.052 to 0.016[g]	-\$1.05 to \$1.02
Loomis & White (1996)	Endangered Species Values	-0.05 to -1.89[h]	-\$1.05 to -\$1.89

[a]Based on Model One (OLS unrestricted), Model Two (OLS restricted), and Model Four (multilevel-random effects). Trend parameter was not statistically significant in Model Three (WLS) and Model Five (2SLS).
[b]Based on Model 1 (semi-log unweighted), Model 2 (trans-log unweighted) and Model 3 (semi-log weighted).
[c]Based on a model excluding judgemental variables and a fully specified model, respectively.
[d]Based on an optimized model and a fully specified model, respectively.
[e]Based on Model One (unrestricted) and Model Four (weighted, unrestricted) for stated preference study year. Model Two (methodology only) was not significant for stated preference study, and none of the parameters for trends in travel cost study or random utility model study were statistically significant.
[f]Based on a multilevel (random effects) model and conventional (OLS) model, respectively.
[g]Trend parameter was not statistically significant in three models with different specifications.
[h]Based on a double-log model and linear model, respectively.

TABLE 2
REGRESSION-BASED DUMMY VARIABLE TESTS OF PUBLICATION BIAS

Source	Resource	Unit of Analysis	Significance	Direction of published to unpublished[a]
Smith & Huang (1995)	Air quality	WTP via hedonic property method	Significant	<
Woodward & Wui (2001)	Wetlands	WTP via various methods	Insignificant	<
Zelmer (2003)	Public goods	Voluntary contributions	Insignificant	<
Rosenberger & Stanley (2006)	Recreation	WTP via various methods	Significant	<
Van Kooten (2004)	Carbon sequestration costs in forests	Cost	Significant	>
Dalhuisen et al. (2003)	Residential water demand	Price elasticity	Significant	<[b]
		Income elasticity	Significant	<[c]
Gallett & List (2003)	Cigarette demand	Price elasticity	Significant	>[d]
		Income elasticity	Significant	>[e]

NOTE: This table is adapted from Rosenberger and Stanley (2006).
[a]Gallett & List (2003) created a dummy variable identifying estimates published in the top 36 premier journal. Dalhuisen et al. (2003) created a dummy variable identifying unpublished estimates. Van Kooten (2004) created a dummy variable identifying estimates published in peer-reviewed sources.
[b]Smaller absolute values for price elasticities in unpublished studies.
[c]Greater absolute values for income elasticities in unpublished studies.
[d]Greater absolute values for price elasticities in top journal publications.
[e]Greater absolute values for income elasticities in top journal publications.

Spatial Limits of the TCM revisited: Island Effects

John B. Loomis[*], Juan Marcos González[*‡],
Armando González-Cabán[†]
[*]Department of Agricultural and Resource Economics, Colorado State University
[†]USDA Forest Service, Forest Fire Laboratory, Riverside, California
‡Corresponding Author
Phone: +1-970-491-6325
Fax: +1-970-491-2067
Email: jmarcos@colostate.edu

Selected Paper prepared for presentation at the Western Agricultural Economics Association Annual Meeting, Portland, OR, July 29-August 1, 2007

Abstract

The purpose of this paper is to address a problem that may arise with the assumption of a continuous spatial market in the TCM model. We find that this assumption can be challenged by geographical limitations that an area of study might have. Particularly for islands (or isolated island-like areas) that have a valuable non-market resource or good, the spatial market characteristic of the TCM model might be limited or truncated. The geographical truncation limits the observed maximum travel cost of the demand curve falsely implying a lower WTP than otherwise. The study uses a dichotomous choice CVM to confirm that the resulting demand schedules from the TCM underestimates WTP for day trips to the Caribbean National Forest in Puerto Rico. This results in a considerably smaller TCM WTP for the value of recreation sites at \$17 to \$29 versus \$109 per day trip from the dichotomous choice CVM. .

JEL Classifications: Q0 Agricultural and Natural Resource Economics

Key Words: Contingent Valuation, count data, consumer surplus, Puerto Rico river recreation, travel cost models.

Introduction

The ideas behind the Travel Cost Model (TCM) were first suggested by Harold Hotelling in 1949 and later on extended to recreation by Marion Clawson. The model recognizes that recreation sites, even when people did not pay entrance fees, have an implicit price that stems from the costs involved with visiting the site. This *travel cost* includes both travel cost and travel time to get to the site. The idea of using an implicit price served to develop a demand-based model (analog to those commonly used in regular goods' demand) that could be used to value recreational uses of the environment (Parsons, 2003). Implicitly then, the TCM also relies upon the notion of a spatial market where visitors' willingness to trade travel costs for site visits reveals their willingness to pay (WTP) for the site and its characteristics. By looking across people who live at different distance from the recreation site hence face different travel costs, the model allows researchers to estimate a "revealed" demand curve for a site and its components.

Determining the travel cost incurred by each visitor has been one of the most researched aspects in the TCM literature. These efforts include studies that look at the opportunity cost of time (Larson and Shaikh, 2001), latent separability of costs (Blundell and Robin, 2000) and how to separate on-site time from travel time (Shaw, 1992; McConnell, 1992). In addition, past research has focused on the assumptions of the TCM that distant visitors actually incur the travel cost exclusively to visit the site of interest (the so-called multiple destination trip bias problem)(Haspel and Johnson, 1982; Mendelsohn et al., 1992), but very little research has focused on physical or natural spatial limits to the travel cost model. The closest concern in using TCM is in urban recreation settings where there may be insufficient variation in travel costs to fully reflect a visitor's WTP (Loomis and Walsh, 1997).

A similar, but somewhat different problem arises in the case of recreation that take place on small islands such as Hawaii, Puerto Rico, Jamaica etc., i.e., islands with significant resident populations that visit local sites. The difficulty on these islands is the maximum travel cost that a visitor can incur is limited or truncated by the physical size of the island. If the site is of high value to the locals, such that their maximum WTP exceeds the maximum cost associated to the distance necessary to drive, this will not be reflected in a typically estimated trip frequency model (e.g., count data model of recreation). That is, the choke price may be constrained below the maximum WTP by the physical distance of the island. In this case, TCM will under-estimate visitors maximum WTP because it appears to the model that visitation stops at this physically imposed choke price, and there is no consumer surplus, i.e., WTP beyond this level. This is particularly a problem with on-site sampling in which we only observe visitors, that is people who even at the highest observed travel cost still take one or more single destination trips. With on-site sampling we cannot observe the zeros.

In our data from Puerto Rican residents visiting streams on the Caribbean National Forest, the maximum observed travel cost was approximately $60 (strongly influenced by the 100 mile width of the island). To allow respondents WTP to not be constrained by this physical limit on the choke price, we asked them if they would still take their most recent trip at a random **increase** in the bid amount that was upwards of $200. This additional question allowed us to look at the same valuation problem from a

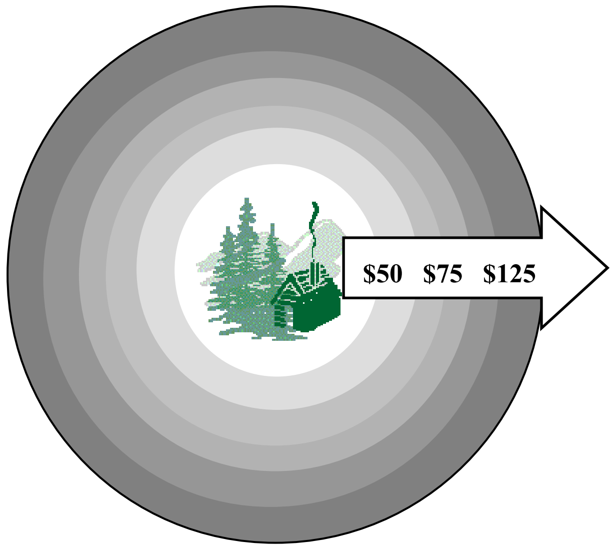
CVM perspective and proves useful as it shows how much the TCM under estimates people's WTP.

In the next sections we elaborate on the idea of truncated spatial markets and how this can affect the WTP measures that researchers obtain when using TCM. Then, we discuss the empirical application in which this truncation is seemingly observed, explain the methodology followed to determine individual's WTP under each type of model and present the results obtained from them. Finally, we look at future areas of research in this area.

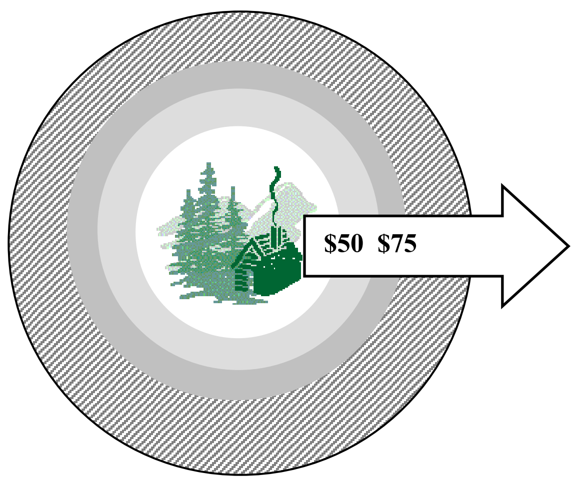
A Truncated Spatial Market

The TCM assumes that people from different points can travel to a given site. Because a main component of the implicit price in the model has to do with time traveled, travel cost is understood to increase in a continuous fashion as one gets further away from the site of interest. Figure 1.A. shows a representation of this spatial property of the travel cost. In the representation one can see that the cost of visiting a site increases as we move to the outer rings of the diagram. On the other hand, figure 1.B. shows what would happen if the spatial market was truncated and the geographical area around the site was limited. In this case, the maximum amount observed is lower than the one we see in diagram A. Even if the site was worth more to the average person in the inner rings, they would not have the chance to reveal it because they have no need to do so. In essence, the demand curve is truncated at the maximum amount of money needed to visit the site from any particular point of the island.

Figure 1. A) Continuous Spatial Market Assumed by TCM and

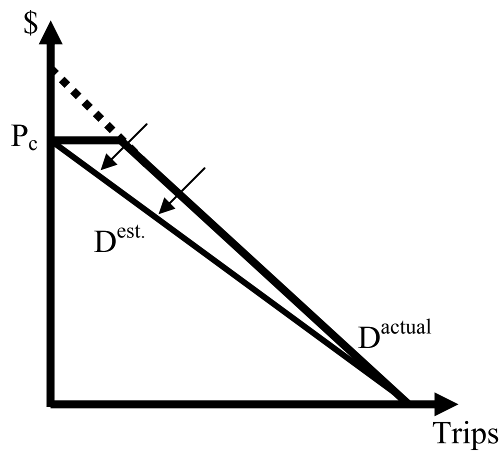


B) Example of Truncated Spatial Market



As presented in figure 2, the reduction in WTP (hence consumer surplus) caused by spatial truncation can come from two different sources. First, when calculating consumer surplus from visitors' revealed preferences, the researcher does not observe any portion of the demand curve that is above the choke price P_c. The area above this price is not revealed to the researcher, thus it cannot be accounted for despite being a real gain for consumers. Furthermore, because TCM valuation studies make use of fully parametric regressions (count data models), the demand curve estimated by them adjusts itself to the information it has, tilting the schedule down towards the choke price.

Figure 2. Truncated Demand Schedule



As a result of this the estimated demand curve ($D^{est.}$) appears flatter than the actual demand schedule (D^{actual}). Not only would the researcher miss the portion of the demand that is above the truncated price level, but it would also force the estimated demand to adjust to this lack of information beyond P_c and cause a further "loss" in consumer surplus.

Methodology

To measure the degree of under-estimation in visitors WTP from the TCM in a constrained island environment, we compare our TCM estimates to those estimated from a dichotomous choice Contingent Valuation Method (CVM). CVM does not suffer from the physical limits as it increases the travel cost by a random amount so a difference between the two WTP measures **could be** attributed to the situation explained above.

Likely, any difference between TCM and CVM estimations is not due to hypothetical bias or other biases associated with CVM. In 1996 Carson et al. used over

600 different CVM and TCM estimates and concluded that differences between CVM and TCM WTP were not statistically significant. If any, CVM WTP measures are generally below TCM WTP estimates (roughly .9 of TCM estimates).

In the TCM case, we use a traditional count data model. To account for possible overdispersion a negative binomial distribution was chosen and robust standard errors were obtained for each coefficient in the specified model. Two set of parameters were estimated under the TCM. The first one uses the on-site correction described by Englin and Shonkwiler (1995). However, on-site WTP values are smaller than the uncorrected WTP values because they are meant to obtain the surplus of the general population not just the visiting portion. With this in mind, the study also looks at the uncorrected TCM equivalent so both visitor groups can be compared. For the dichotomous choice CVM a probit distribution was chosen. In both models (CVM and TCM) the observations considered were limited to those where individuals who indicated that visiting the site was the main purpose of their visit. This was done to control for the possible multiple destination problem mentioned before and found sometimes in on-site samples.

Once the coefficients for the models are obtained mean WTP measures are calculated following TCM and CVM theory and considering the distributional assumptions made. An empirical convolution process follows in order to statistically determine whether differences in WTP measures are significant. The method proposed by Poe et al. in 2005 is intended to find all possible differences between two sets of values. By exploiting the distributional assumptions about the model parameters we generate a random vector of WTP values within the coefficients' confidence. The convolutions method then looks at these vectors and determines the probability that one WTP distribution lies on top of the other. The resulting p-values are then used as statistical ground to test that CVM and TCM WTP measures are indeed different.

Empirical Application

The study uses data set from a survey administered in the Caribbean National Forest in Puerto Rico. The on-site surveys contain information on trip demand for the 2005 season and a CVM question that was meant to complement the trip assessment. Data were collected at 11 different sites within the forest and contained demographic information of the users, distance and time traveled, and characteristics of the visited sites.

Over 700 observations were obtained and coded, of which 430 observations were used in this analysis. The reason for the reduction in observations is because only trips where visiting the site were the main reason for traveling are considered valid for the TCM. This is done to deal with multiple destination problems (274 trips were not single destination trips). As mentioned before, these observations are typically pointed out as a source of distortion in travel cost models. Also, because of the complicated form of the corrected negative binomial distribution, we eliminated visitors who took more than 100 trips because they appear to be from visitors that are somehow quite different than the vast majority who take a small fraction of these trips.

Variables in the models include an **intercept**, **travel cost (**in the TCM case) and a **bid** amount visitors were asked to pay (in the CVM case). The model also includes **road** (as a measure of accessibility), **mean annual stream discharge** (as a measure of average seasonal flow)**, distance of pool to bridge, pool volume, streamflow day** (as a measure of flow during visit), the number of **picnic tables** at the site and **median grain size** (measure of substrate sand size). A dummy was also included to indicate whether the site had a **waterfall**, and whether there were **formal trails** and **restaurants** in the area of interest. Finally a dummy variable was also used to define whether the visitor was male or female. Separate regressions indicate these variables have the greatest explanatory power under each model. The following is a table that presents the summary statistics for the variables used.

Table1. Summary Statistics

	Mean	Minimum	Maximum
Bid	64.02196	1	200
Travel Cost (TC)	7.942791	0.259804	68.72794
Road	3.607921	2	5
Mean Annual Discharge	0.82763	0.106	1.667
Dist. Pool to Bridge	23.84158	0	145
Median Grain Size	462.5208	102	2337
Pool Volume	460.2487	42	1868.4
Gender	0.524851	0	1
Waterfall	0.479125	0	1
Streamflow Day	39.37861	9.2	108
Picnic Tables	0.544304	0	3
Trash Cans	4.784	0	13
Formal Trails	0.489109	0	1
Restaurants	0.135354	0	1

Results

Three models were used for the purpose of this study. The results of these models are summarized in Table 2. In all cases, the values obtained in the regression follows what theory suggests with a negative and significant bid and travel cost coefficient. These yielded a $17 WTP for the corrected TCM, $29 for the uncorrected version of it and $109 for the CVM. It should be mentioned that the highly significant value for alpha in the TCM results suggests we correctly chose a negative binomial distribution. As expected, the WTP measures for the corrected negative binomial distribution are lower than the uncorrected version and both TCM WTP values are well below the CVM analog.

Table 2. Results from parametric regressions using CVM and TCM models.

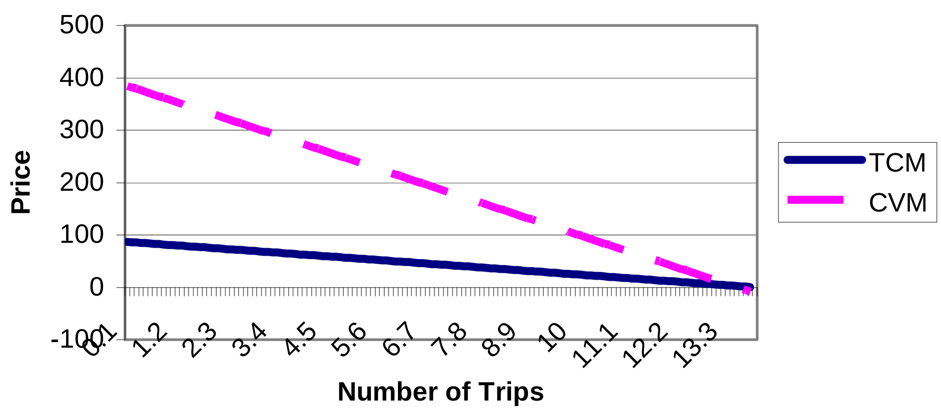
Variable	CVM Coef. (Std. Error)		TCM (Corrected) Coef. (Std. Error)		TCM Coef. (Std. Error)	
Bid/TC	-0.0104	***	-0.0576	***	-0.0343	***
	(0.001149)		(0.0175525)		(0.008647)	
Road	-0.2485	**	0.1508	*	0.1323	**
	(0.10296)		(0.0825145)		(0.063739)	
Mean Annual Discharge	-0.5113	*				
	(0.304429)					
Dist. Pool to Bridge	0.0012					
	(0.002557)					
Median Grain Size	-0.0003	**				
	(0.000169)					
Pool Volume	0.0004	*				
	(0.000249)					
Gender	0.1846					
	(0.128021)					
Waterfall			0.3394		0.2455	
			(0.2473462)		(0.202802)	
Streamflow Day			-0.0042		-0.0033	
			(0.0052275)		(0.004084)	
Picnic Tables			-0.6497	***	-0.3489	***
			(0.1769431)		(0.118958)	
Trash Cans			0.0563		0.0303	
			(0.0591204)		(0.042084)	
Formal Trails			-0.4654	*	-0.3876	*
			(0.256722)		(0.203329)	
Restaurants			0.6965	*	0.5263	*
			(0.3606127)		(0.297835)	
	2.2962	***	-15.5405	***	1.4616	***
	(0.584221)		(0.4559102)		(0.36982)	
/LN(alpha)			16.7613	***		
			(0.146858)			
alpha					1.0105	***
					(0.073504)	
Pseudo Log Likelihood	**-260.3699**		**-1013.9264**		**-1139.0408**	
Mean WTP	**$ 109.48**		**$ 17.37**		**$ 29.16**	

Significant at the 90% confidence level, ** significant at the 95% confidence level, *** significant at
the 99% confidence level.

Results from the empirical convolutions show that in both cases (corrected and uncorrected) the CVM WTP is statistically different from the TCM WTP measures. A two tail p-value of 0.0053 and 0.0019 for the comparison between CVM WTP and the uncorrected and corrected TCM respectively showed that neither TCM WTP distributions overlaps the CVM WTP. This is not surprising considering the WTP obtained for the dichotomous choice CVM is 3.6 times greater than the uncorrected TCM WTP and more than 6 times greater the WTP obtained from the corrected TCM.

Figure 3. shows that the effect of the island's physical size limit determining the choke price in the "continuous" count data model also biases the slope coefficient. So the reduced WTP with the TCM is a combination of the censored choke price and its effect on the price coefficient. Figure 3 also illustrates what the implied demand curve from the CVM looks like.

Figure 3. Implied Demand Curves for Recreational Trips Under CVM and TCM



Conclusions

The count data TCM corrected for on-site sampling bias had a negative and statistically significant travel cost coefficient. This yielded an average net WTP $17 per trip. The dichotomous choice CVM had a negative statistically significant bid coefficient. The CVM yielded an average net WTP of $109 per trip. As can be seen this is a sizeable difference given that both are modeling the exactly the same people at the same sites. Our interpretation is that the higher WTP estimate from the dichotomous choice CVM is more reflective of the high quality visitor experience and the visitors' net WTP than would be the TCM.

Our very large difference in net WTP per trip is due to the physical size limit of the island of Puerto Rico. It would be interesting to repeat this type of TCM and CVM analysis at similar quality recreation sites on islands of different sizes to see what the relationship is. As an island grows in size relative to the quality of the recreation site, the difference in the WTP estimates should be less pronounced. Alternatively, on islands smaller than Puerto Rico the bias could even be much larger. Researchers need to be aware of this concern when doing local recreation site valuation on islands where most of the visitor use is by island residents.

Future research could also focus on using simulations to look at what happens to the estimated demand schedule in the TCM as truncation is eliminated by gradually expanding the population to a complete and continuous spatial market. This should provide relevant evidence to further identify the limits of the TCM and this particular geographical assumption.

References

Blundell, R. and J.M. Robin, 2000. Latent Separability: Grouping Goods Without Weak Separability. Econometrica, 168 (Jan), 53-84.

Carson, Richard T., N. Flores, K.M. Martin and J.L. Wright 1996. Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods. Land Economics, February v. 72, iss. 1, pp. 80-99

Englin, J., Shonkwiler, J. S., 1995. Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation. Review of Economics and Statistics. 77, 104-112.

Haspel, A.E. and F.R. Johnson, 1982. Multiple Destination Trip Bias in Recreation Benefit Estimation. Land Economics 58 (Aug), 364-372.

Larson, D. and S. Shikh, 2001. Empirical Specification Considerations for Two-Constraint Models of Recreation Demand. American Journal of Agricultural Economics 83 (May): 428-440.

Loomis, John B., and Richard Walsh. 1997. Recreation Economic Decisions: Comparing Benefits and Costs. State College, Pennsylvania: Venture Publishing, Inc.

McConnell, K., 1992, On-Site Time in the Demand for Recreation. American Journal of Agricultural Economics 74 (4): 918-924.

Mendelsohn, R.J. Hof, G. Peterson and R. Johnson, 1992. Measuring Receation Values with Multiple-Destination Trips. American Journal of Agricultural Economics 74, 926-933

Parsons, G.R., 2003. The Travel Cost Model, in: Champ, P., Boyle, K.J., Brown, T.C. (Eds.), A primer on Nonmarket Valuation, Vol. 3. Kluwer Academic Publishers, Netherlands, pp. 269-329.

Poe, G.L., Giraud, K.L. and J.B. Loomis. 2005. "Computational Methods for Measuring the Difference of Empirical Distributions." American Journal of Agricultural Economics, 87(2): 353-65.

Shaw, W.D., 1988. On-Site Samples Regression: Problems of Non negative Integers, Truncation and Endogenous Stratification. Journal of Econometrics 37, 211-223.

Shaw, W.D. 1992. Opportunity Cost of Time. Land Economics 68 (1), 107-115

A CONJOINT ANALYSIS OF LOCAL STAKEHOLDER VALUES FOR TROPICAL PROTECTED AREAS

Donald F. Dennis
Research Forester, USDA Forest Service
Northern Research Station
705 Spear Street
South Burlington, VT 05403
(802)951-6771 ext.1080, Fax (802)951-6368
email: dondennis@fs.fed.us

Michel Masozera
University of Vermont

Jon Erikson
University of Vermont

Abstract: The continued delivery of ecosystem services produced in tropical areas is essential to economic prosperity and human welfare. Because local people are on site, and often have an intimate relationship and dependence on the land their input and support can be crucial to the success of any strategy for protecting tropical areas. This study uses conjoint analytic techniques to assess and analyze local stakeholder values with respect to establishment and regulation of 2 large protected areas in Madagascar. The primary focus is on survey design and analytical methods. Preliminary analyses indicate that watershed protection was the most important attribute for local stakeholders and that there are regional differences in the preferences of local stakeholders.

INTRODUCTION

Natural ecosystems and the biological diversity contained within them provide a range of goods and services that include, among many others, biodiversity conservation, carbon sequestration, watershed protection, soil formation, and scenic beauty. The continued delivery of these services is essential to economic prosperity and human welfare. Madagascar is one of 18 recognized mega diversity countries that represent 80% of the world's biodiversity and is a high priority for international conservation efforts. The majority of the biodiversity is found in the forests, which are also the direct source of livelihood for over 90% of the country's population. An annual population growth rate of 3% and a subsistence dependency on slash and burn agriculture pressures the remaining forests. As in other developing countries, the standard approach to slow ecosystem degradation has been to establish protected areas and regulation access. Effectiveness has been limited because establishment of protected areas often takes place at considerable social and economic costs to local communities in terms of access to land, timber, wildlife, and other resources.

Effective management for the provision of ecosystem services must consider the balance of tradeoffs over varying geographic and institutional scales (Hein et al. 2005). Local stakeholders may be asked to bear the burden of lost opportunity costs for protected resources while receiving little benefit. For example, carbon is sequestered at the local level while the benefits are primarily global. Stakeholders at different scales may have different perspectives on the values of ecosystem services based on cultural background, dependency on resources for their livelihood, and other socio-economic characteristics. This often leads to different visions for management and conflicts of interest. Because local people are on site, and often have an intimate relationship and dependence on the land their input and support can be crucial to the success of any protection strategy. There is a very real threat that protected areas can become "paper" parks with protection falling far below the mandated standards.

This study uses conjoint analytic techniques to assess and analyze local stakeholder values with respect to establishment and regulation of 2 large protected areas in Madagascar. The Masoala National Park contains 230,000 ha. of rainforest and 10,000 ha of marine park, and the Makira Conservation Site is 450,000 ha. of tropical forest with many lake and river habitats. Conjoint techniques are well suited for soliciting and analyzing preferences in environmental decisions that frequently entail tradeoffs between costs and benefits that are not represented efficiently in market transactions (Dennis, 1998). We used a conjoint ranking survey to elicit local stakeholder values and acceptable tradeoffs for varying levels of 4 attributes that are related to the protected areas: watershed protection (water quality and quantity), wildlife habitats, availability of opportunities for recreation and ecotourism, and the type and extent of the protection or conservation strategy employed for the protected areas. See Table 1 for a description of the attributes and levels used in the study. Both linear and quadratic effects were estimated.

Table 1. Attributes and levels.

Watershed protection (water quality and quantity).
1. Water quality for irrigation is insufficient and water quality for drinking and downstream fish habitats is bad most of the year over the next 15 years.
2. Two or three month (per year) shortage in water supply for irrigation and water quality for drinking and downstream fish habitats is good all year over the next 15 years.
3. Water supply for irrigation is guaranteed and water quality for drinking and downstream fish habitats is better all year long over the next 15 years.

Wildlife habitat.
1. Deterioration of wildlife habitat that protects rare and endangered species.
2. Maintain the current level of improvement in wildlife habitat that protects rare and endangered species.
3. High improvement of wildlife habitat that protects rare and endangered species.

Recreation/ecotourism.
1. Deterioration/decrease from the current level of tourism revenue.
2. Maintain the current level of increase in tourism revenue (10% per year) over the next 15 years.
3. 15% increase in tourism revenue per year over the next 15 years.

Type of protection/conservation strategy.
1. Limited access to forest resources through "transfert de gestion"[1] to COBA[2] but with no government supervision/regulation.
2. Limited access to forest resources through "transfert de gestion" [1] to COBA[2] but with government supervision/regulation.
3. No access/strict government control.

[1] "transfert de gestion" is to transfer management of public forest to local communities. There is new legislation that gives the government authority to enter into contractual arrangements with communities for land management.

[2] COBA stands for "communaute de base" in French which is a community level forest association. The COBA is made up local forest users, primarily residents who use forests for firewood, timber, medicinal plants, food, and cultural practices. To be granted a contract, a COBA must have official standing as an association and be sanctioned by the mayor's office.

Although preliminary results are reported and discussed, the primary focus of this paper is on an overview of the problem, survey design, development of the conjoint or choice model, and analytical capabilities. Future analyses will examine the marginal rates of substitution (tradeoffs) among the attributes at differing values, as well as regional and demographic differences in preferences among stakeholders.

METHODS

Conjoint analysis, a form of choice modeling, is a technique for measuring psychological judgments that is used frequently in marketing research to measure consumer preferences for products with multiple attributes (Green et al. 1988). Respondents choose between alternative products or scenarios that display varying levels of selected attributes. The utility or preference for each attribute can be inferred from the respondent's overall evaluations. These partial utilities, or partworths, indicate the relative importance of each attribute's contribution to overall preference or utility. They can be combined to estimate relative preferences for any combination of attribute levels.

Analytical Model

A random utility model is used to explain local stakeholder preferences toward various environmental, economic, and social aspects associated with designation and management of protected areas. When presented with a set of alternatives, individuals are assumed to make choices that maximize their utility or satisfaction. The utility that the i^{th} individual derives from the j^{th} alternative (U_{ij}) can be represented as:

$$U_{ij} = X'_{ij}\, \beta + e_{ij} \tag{1}$$

where X_{ij} is a vector of variables, which may include transformations of variables, that represent values for each attribute of the j^{th} alternative to the i^{th} individual; β is a vector of unknown parameters; and e_{ij} is a random disturbance, which may reflect unobserved attributes of the alternatives, random choice behavior, or measurement error. In the empirical study under consideration, a respondent's utility level (U_{ij}) for each alternative is not observed, but a ranking (r_j) is observed that corresponds to the order of his or her underlying utilities. The probability of alternative 1 being ranked above other alternatives is:

$$\begin{aligned} P_{i1} &= \Pr(U_{i1} > U_{i2} \text{ and } U_{i1} > U_{i3} \text{ ... and } U_{i1} > U_{iJ}) \\ &= \Pr[(e_{i2} - e_{i1}) < (X'_{i1}\,\beta - X'_{i2}\,\beta) \text{ ... and } (e_{iJ} - e_{i1}) < (X'_{i1}\,\beta - X'_{iJ}\,\beta)] \end{aligned} \tag{2}$$

Similar expressions hold for each of the remaining alternatives being chosen next in the choice set, and the P_{ij}'s become well-defined probabilities once a joint density function is chosen for the e_{ij} (Judge et al. 1985).

McKelvey and Zavoina (1975) developed a polychotomous probit model to analyze ordinal level dependent variables. They assume that the e_{ij}'s are distributed normally with mean 0 (the variance is standardized to unity), and that the observed variable (Y_{ij}, the ranks for the J alternatives) is related to the true unobserved utilities (U_{ij}) in the following way:

$$Y_{ij} = 0 \text{ if } U_{ij} \le \mu_{i1},\ Y_{ij} = 1 \text{ if } \mu_{i1} < U_{ij} \le \mu_{i2}, \text{ ... } Y_{ij} = J\text{-}1 \text{ if } U_{ij} > \mu_{iJ-1}. \tag{3}$$

The μ_{ik}'s define the boundaries of the intervals for the unobserved utilities that correspond to the observed ordinal response. Since the μ's are free parameters, there is no significance to the unit distance between the set of observed values of Y; they merely provide the ranking.

Estimates are obtained by maximum likelihood and the probabilities entering the log-likelihood function are the probabilities that the observed ranks (Y_{ij}'s) fall within the J ranges defined by J+1 μ's. The parameters to be estimated are J-2 μ's plus the β vector; μ_0 and μ_J are assumed to be negative and positive infinity, respectively. McKelvey and Zavoina (1975) describe the model and maximum likelihood estimators in greater detail.

In the polychotomous probit model the estimated value ($X'_{ij}\,\beta$) for an observation determines the position of the mean of the distribution of response categories over the underlying scale. The μ's delineate ranges in the unobserved underlying variable (utility) that correspond to the observed response categories. The estimated probability that a response will fall in each category or rank in the case under consideration is measured by the area under the normal standard density curve $[f(X'_{ij}\ \beta)]$ and bounded by the respective μ's. These probabilities can be computed using the estimated model parameters:

$$\Pr(Y_j = k\text{-}1) = \Pr(U_j \text{ is in the kth range}) = F(\mu_k - X'_j\ \beta) - F(\mu_{k-1} - X'_j\ \beta) \qquad (4)$$

where k indexes the rankings and F(•) is the cumulative standard density function, assumed normal for the probit specification. Thus, the effect of a discrete change in the level of the nth independent variable (x_{nj}) on the estimated probability that a response will fall within each of the categories (ranks) can be calculated by substituting the estimated parameters (β and μ's) into Equation 4. In probit analysis the estimated coefficients (β) represent the effect of a unit change in an independent variable on the underlying scale given by $X'_j\ \beta$. Graphically, this is shown as a shift in the distribution of responses over the underlying scale. The magnitude of that change on the probability of a particular response occurring will depend on the original position as determined by all the estimated parameters and associated variables.

Survey Design and Administration

A conjoint ranking survey was used to assess stakeholder values. The survey was translated into Malagasy and pretested with park staff from diverse backgrounds. Based on this pretesting the survey was revised to minimize technical terms and reduce the length of descriptions. The revised survey was administered to individual stakeholders within a group meeting setting. With help from the Masoala National Park and Makira Conservation Site staffs four villages were selected in which to hold meetings and potential stakeholders to invite were identified. Twenty-five people representing different socioeconomic and demographic groups from each village were invited to participate in the meetings. Invitations were sent out by park managers and village leaders with care taken to include as many stakeholder groups as possible (e.g. farmers, fishers, men, women, youths, etc.). In each meeting facilitators gave an overview of the nature and purpose of the survey and a detailed explanation of the choice attributes and levels. Participants had the opportunity to ask questions or discuss their concerns. Respondents who could not read or write were assisted in a manner that would not influence their choices.

Each respondent ranked 9 alternatives, each with 1 level for each of 4 attributes. The combinations of attribute levels for the alternatives were chosen using an orthogonal design that allows estimation of linear and quadratic main effects over the entire 81 (3^4) possible attribute combinations, with the least number of choice responses. Table 1 lists the attribute levels and Figure 1 shows a sample alternative. Respondents also completed a series of socioeconomic and demographic questions.

Figure 1. A sample alternative.

Alternative #1 -------------------- ***Packet #-------***

Water supply for irrigation is guaranteed and water quality for drinking and downstream fish habitats is better all year long over the next 15 years.

Deterioration of wildlife habitat that protects rare and endangered species

Maintain the current level of increase in tourism revenue (10% per year) over the next 15 years

No access/Strict government control

RANK ---------

An ordered probit model (described in section 2.1) was used to analyze the responses to the ranking survey. The dependent variable is the ordinal ranking of the alternatives which is coded 0 to 9; higher scores being associated with greater utility. The independent variables (attribute levels 1, 2, 3 in Table 1) were coded, respectively, -1, 0, 1 for the linear form and 1, -2, 1 for the quadratic form. This coding scheme maintains the ordinal relationship for the linear term and provides for an orthogonal contrast with the quadratic term.

RESULTS AND DISCUSSION

The survey was given to 87 stakeholders living in the Masoala and Makira regions. Each respondent ranked 9 alternative scenarios for a total of 783 preference rankings. Preliminary analyses indicate that all linear effects were significant at the 5% level and the quadratic effect was significant only for the recreation/ecotourism attribute (Table 2). The relative importance scores shown in Figure 2 were computed by dividing the utility range for each attribute by the sum of the utility ranges for all attributes. These scores indicate how important a particular attribute was in the overall preference for alternatives but not whether changes in the level of the attribute had a positive or negative influence on preference. The signs and magnitude of the estimated coefficients or partworths supply that information.

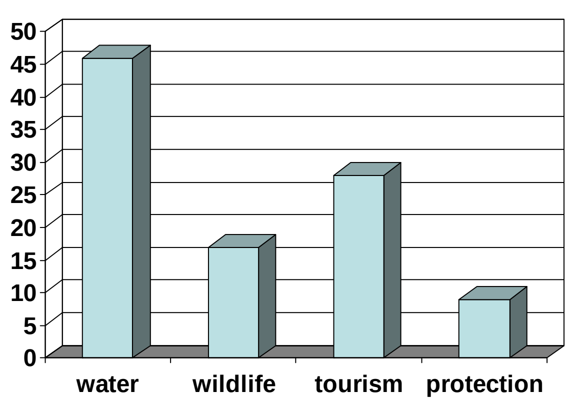
Table 2. Parameter estimates for an ordered probit model.

Variable	Coefficient	Std. Error	Pr>ChiSq
Linear effects:			
Water	0.717	0.0487	<.0001
Wildlife	0.405	0.0793	<.0001

Tourism	0.267	0.0638	<.0001
Protection	-0.133	0.0646	0.0391
Quadratic effects:			
Water	-0.003	0.0261	0.9136
Wildlife	0.105	0.0693	0.1305
Tourism	0.202	0.0528	0.0001
Protection	-0.063	0.0518	0.2269
Boundary parameters:			
μ_1	-1.5600		
μ_2	-0.9518		
μ_3	-0.5143		
μ_4	-0.1374		
μ_5	0.2227		
μ_6	0.5792		
μ_7	0.9792		
μ_8	1.4889		

Log likelihood = -1561.
N = 783 (87 respondents, 9 preference rankings each)

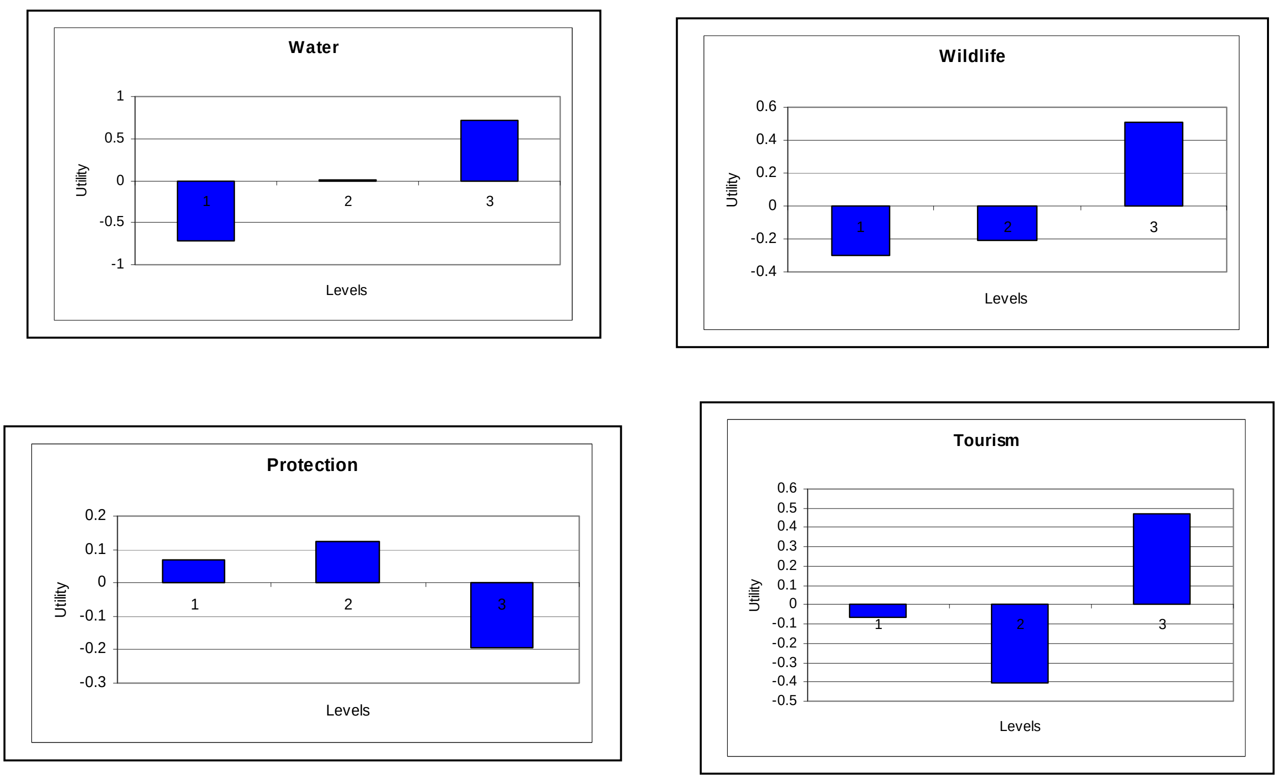
Figure 2. Relative importance scores.



Watershed protection was the most important attribute for local stakeholders. Water is important for domestic use in both regions. In the Makira region respondents indicted that improved water supply would enhance prospects for agricultural irrigation while those in the Masoala region cited enhanced fisheries as a benefit of an improved water supply.

Changes in opportunities for recreation and tourism was the second most important attribute. The significance and magnitude of the quadratic effect indicates that respondents preferred to either allow tourism to deteriorate or to enhance possibilities for tourism over maintaining the current level of increase in tourism (Figure 3). Verbal comments and discussions indicated that they felt this way because they see little benefit from the current trend.

Figure 3. Partial utilities (partworths) for each attribute level.



Stakeholders clearly preferred a high improvement in wildlife habitat (level 3) but showed only a slight preference between levels 1 and 2. Apparently they were not satisfied with the current trends.

Local stakeholders were against loosing all access to protected areas and strict government control (level 3). Beyond this basic view, initial analyses indicated that there was a marked difference between stakeholders in the two study regions. The Masoala National Park has been protected for many years and the local people feel that they have seen little economic benefit. In this region stakeholders preferred the least amount of government control to their access (level 1). But in the Makira region, which is not yet a designated National Park, local respondents were more willing to accept government supervision of access (level 2) with the hopes that they will see improved economic and ecological benefits as a result of protection.

CONCLUSIONS

An understanding and consideration of the values of local stakeholders is deemed important to the ultimate success of protected area management. A conjoint model was given to local stakeholders adjacent to protected areas in Madagascar. Preliminary results indicate a hierarchy of importance in the various attributes associated with protected area management. Water quality and quantity was clearly of the greatest concern to local stakeholders. They were also dissatisfied with current trends in wildlife management and the economic benefits they receive from ecotourism. Although stakeholders in both regions were against loosing their access to the protected areas entirely and strict government control, stakeholders in the Makira region were more willing to accept some government regulation and limits to their access in the hopes of seeing improvements in the quality of their lives from the ecological (primarily improved watersheds) and perceived economic benefits that might result from protected areas.

REFERENCES

Dennis, Donald F. 1998. Analyzing public inputs to multiple objective decisions on National Forests using conjoint analysis. Forest Science, 44(3):421-429.

Green, P.E.; Tull, C.S.; Albaum, G. 1988. Research for Marketing Decisions. 5th ed. Englewood Cliffs, NJ, Prentice Hall. 784 p.

Hein L.; van Koppen, K.; de Groot, R. ; van Ierland, E.C. 2006. Spatial scales stakeholders and the valuation of ecosystem services. Ecological Economics, 57(2):209-228.

Judge, G.G.; Griffiths, W.E.; Hill, R.C.;Lutkepohl, H.; Lee, T.C. 1985. The Theory and Practice of Econometrics. Ed. 2. New York, Wiley. 1019p.

McKelvey, R.D.; Zavoina, W. 1975. A statistical model for the analysis of ordinal level dependent variables. Journal of Mathematical Sociology, 4:103-120.

Building Wealth through Ownership: Resident-Owned Manufactured Housing Communities in New Hampshire

Charlie French
Extension Associate Professor, Community and Economic Development
University of New Hampshire
Durham, New Hampshire
Charlie.French@unh.edu

Kelly Giraud
Associate Professor, Resource Economics and Development
University of New Hampshire
Durham, New Hampshire
Kelly.Giraud@unh.edu

Sally Ward
Professor, Sociology
University of New Hampshire
Durham, New Hampshire
Sally.Ward@unh.edu

Abstract

Eighty-two resident-owned manufactured housing parks serve over 4,000 New Hampshire families. Despite the popularity of resident-owned parks, one important question remains; do they outperform investor-owned manufactured housing parks from a social and economic standpoint? A research team from UNH set out to answer this question through a comprehensive study that engaged subjects from resident-owned parks and investor-owned parks, as well as officials from seven New Hampshire towns. The research findings suggest that resident-owned manufactured housing parks indeed provide a more affordable housing option for low-income families, as well as an enhanced sense of ownership and an opportunity to build equity.

Introduction

Between 1984 and 2006, the New Hampshire Community Loan Fund – referred to as 'The Loan Fund' – has helped residents from 82 manufactured (mobile) home parks to purchase the land on which their homes are situated. Each of these cooperatively-owned parks, known as Resident-owned Communities (ROC's), formed a self-governing corporation to manage their park. Through this model of resident ownership, residents have gained financial and managerial control of their park and their lives.

While number of studies examine the social and economic benefits of home ownership (Haurin, Deitz, & Weinburg 2003), few studies examine the social and economic benefits of *cooperative* home ownership. In fact, the concept of cooperative (resident) ownership of manufactured home communities is relatively new, with the first ROC having been established in New Hampshire in 1984. Yet, ROC's have already had a significant impact on the state's affordable housing sector. Today, over 4000 New Hampshire families reside in ROC's – more than in any other state.

Due to the dearth of literature examining ROC's as a model of home ownership, the Ford Foundation and 'The Loan Fund' commissioned the Carsey Institute at the University of New Hampshire to conduct an independent evaluation of the social and economic outcomes of resident ownership of manufactured home communities in New Hampshire.

While there are many theories as to why ROC's have proliferated in New Hampshire, this evaluation examines four specific advantages that ROC's are believed to have compared to investor-owned communities (investor-owned communities, or IOC's, are manufactured housing parks where residents rent the land on which their home is situated). Based on preliminary data collected by 'The Loan Fund' (Bradley, 2002), these advantages are:

- Better access to mortgage financing
- Greater stability in housing costs
- More opportunity to build equity
- Enhanced sense of ownership and control

Additionally, the team was charged with providing 'The Loan Fund' with recommendations that would help them to strengthen the resident ownership model in New Hampshire. Provided that there is strong evidence supporting the advantages outlined above, the hope is that the resulting recommendations will also highlight opportunities for Extension Services to provide educational outreach to help build the capacity of existing and potential ROC's, and thereby help promulgate the resident ownership model nationally.

Note: *According to the Manufactured Housing Institute, a manufactured home is constructed in a factory environment and built to federal safety standards known as HUD Code, whereas a mobile home is simply a manufactured home built prior to 1976, before the HUD Code went into effect (2005).*

Background

Home ownership is the main source of equity for most Americans. However, due to rapidly escalating housing costs relative to personal income, an increasing number of Americans cannot afford to purchase a home (Apgar, 2005).

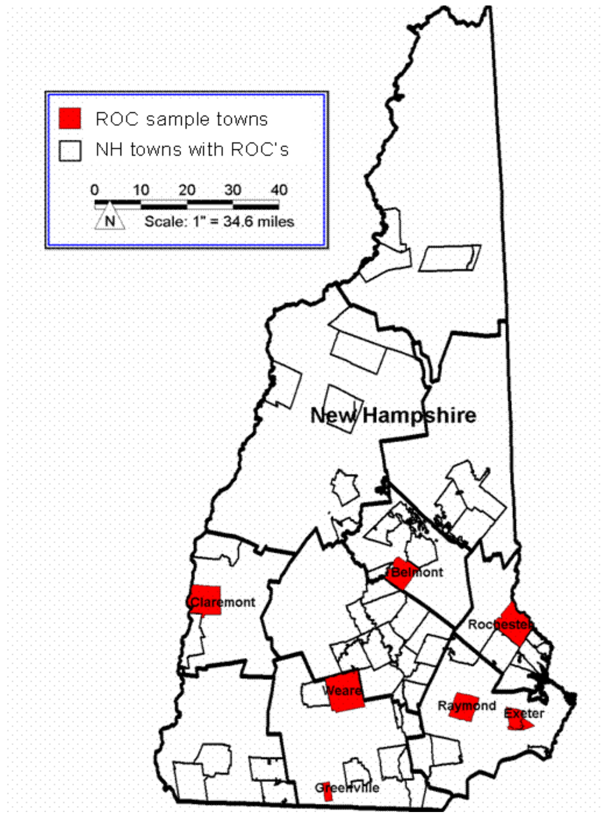
Of those that *were* able to purchase a home in the United States between 1980 and 2000, 29% opted to purchase a manufactured home (Genz, 2001). The vast majority of these manufactured homes are located in IOC's, where residents rent their plot from a landlord. Only a small fraction of manufactured homes are located in ROC's.

A number of factors explain why ROC's have not proliferated nationwide. Foremost, few lenders are willing to provide financing with interest rates comparable to conventional mortgage loans for the purchase of manufactured homes located in parks. Most lenders only provide access to personal property loans or variable-rate loans for the purchase of manufactured homes. Interest rates for these loans are typically several percentage points higher than conventional home loans (Bradley, 2003). And, without access to mortgage-competitive financing, prospective homeowners may not see advantages to buying a home in an ROC. In New Hampshire, however, mortgage-rate financing is increasingly becoming available for the purchase of manufactured homes in ROC's.
Another factor limiting the expansion of ROC's is the lack of technical, financial, and managerial support to the ROC Boards responsible for managing and maintaining their respective parks. New Hampshire is one of the few states where support is readily available to ROC Boards. Without this support, it is unlikely that residents of New Hampshire's 82 ROC's would have been able to purchase their parks in the first place.

Methods

To determine if resident-owned parks pose certain advantages over investor-owned parks, a study was designed to compare ROC's with IOC's on a range of social and economic variables. Overall, seven New Hampshire towns with at least one ROC and one IOC were selected for the study to encompass a wide geographic distribution and a broad range of demographic characteristics (see Figure 1 for map of towns in sample). Within each of these towns, one to two ROC's and an equivalent number of IOC's were selected for comparison. The parks within each town were selected to be comparable in terms of location, size, and demographics of the park residents. The final sample consisted of 8 ROC's and 12 IOC's (Ward, French, & Giraud, 2005).

Figure 1: Towns in study sample



The primary sources of data for this study were a mailed questionnaire, secondary data from town tax cards and the Multiple Listing Service (MLS), and interviews with ROC Board Members and Town Officials.

Surveys:

Using elements from Don Dillman's *Total Design Method* for conducting surveys (1978), a self-administered survey was mailed to residents of both ROC's and IOC's to query them about basic demographic information, household economic factors, as well as their perceptions about living in their park. All of the residents in the sampled parks were mailed surveys, with the exception of one town, where only 50% of the homes were sampled due to the town's large size and the possibility that a full sample could skew the results. Of the 1,187 surveys sent out, 698 were returned for an overall response rate of 59%. Overall, the response rates were very similar among the two groups, with 356 surveys completed by residents of ROC's and 342 completed by residents of IOC's (Ward, French, & Giraud, 2005).

Town Tax Records and Multiple Listing Service Data:

Town Tax Cards were accessed in order to analyze and compare information on assessed value of homes. Likewise, data from the state's Multiple Listing Service (MLS) were used to compare lot rent fees and the number of days on the market for homes sold in ROC's and IOC's.

Interviews:

In-depth, structured interviews were conducted of Board members from 20 ROC's across the seven sample towns to get their perceptions of the benefits and challenges of living in cooperatively-owned parks. Because IOC's do not maintain formal leaders, there was no way to obtain a compatible sampling for IOC's. In spite of this, the interviews of ROC Board members provided valuable insight regarding how each of the ROC's changed since they converted from an IOC to an ROC. The roles of the ROC Board members interviewed included:7 Board Presidents/Chairs

- 2 Vice Chairs
- 4 Treasurers
- 2 Secretaries
- 1 Infrastructure Coordinator
- 4 Members-at-large

In addition, one to two town officials were interviewed in each of the seven sample towns to determine their perceptions of ROC's and IOC's. The twelve town officials that were interviewed fell into one of three categories; safety officer, elected representative, or administration professional (e.g. planner, assessor, etc.).

Findings

Although there were a number of economic variables examined in this study, only four will be addressed in the following section, as they relate to the four advantages that were proposed above. These variables are access to mortgage-competitive financing, stabilization of housing costs, opportunity to build equity, and sense of ownership and control.

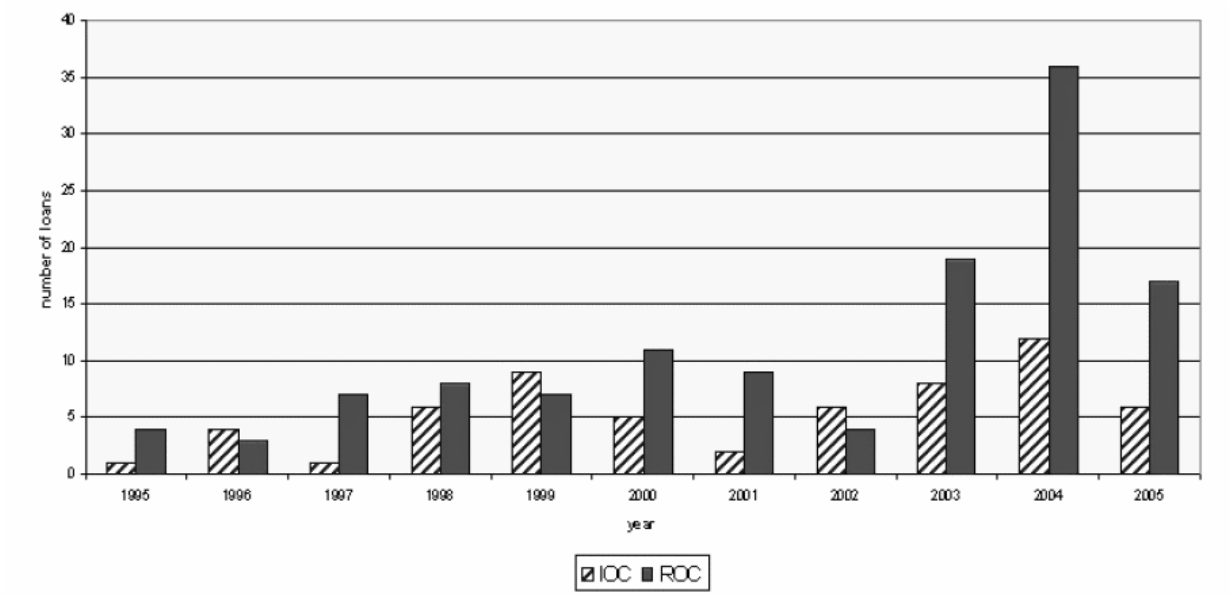
Access to Mortgage-Competitive Financing:

Data on home loans were collected from the surveys to determine whether ROC residents have better access to financing than IOC's. One theme that appeared in numerous surveys was the stated lack of availability of low interest loans for the purchase of manufactured homes.

As Figure 2 below illustrates, homeowners from ROC's obtained more loans to help finance their home purchases than those from IOC's between 1995 and 2005. Moreover, the number of loans to ROC's increased dramatically over the past few years as new lenders made mortgage financing available to homeowners. In contrast, residents of IOC have had fewer loans and were often forced to purchase their homes outright in order to avoid the high interest rates available through personal property loans or variable-rate loans.

Figure 2

Number of Mortgage Loans by Community Type, 1999-2005 (2005 Partial Yr.)



Adapted from Ward, S., French, C., & Giraud, K. (2005). The effect of cooperative ownership on appreciation of manufactured housing. *Cooperative Housing Journal.* 2005/2006 Annual Issue, p. 22.

Stabilization of Housing Costs

The second hypothesis is that ROC's provide more stable housing costs than IOC's. This was tested by comparing monthly lot fees paid by homeowners in ROC's and IOC's. The monthly lot fees paid by homeowners in ROC's are used to pay off their share of the mortgage for the land, as well as for maintenance and improvements. In contrast, monthly lot fees paid by homeowners in IOC's go to the landlord for land rent, much of which is converted to profit.

Table 1
Lot Fee by Community Type

Community type	Summary of monthly lot fee/rent		
	Mean	**Standard deviation**	**Frequency**
IOC	277.62238	36.12557	307
ROC	265.9269	39.524786	342
Total	271.45928	38.373348	649
Anova F prob < .01			

As Table 1 above illustrates, the average lot fee for ROC's ($265.93) is nearly $12.00 per month less than the lot fees for IOC's ($277.62). Taking into consideration that lots in ROC's tend to be larger than lots in IOC's, this appears to be a significant factor. In effect, homeowners in ROC's pay less in lot fees in spite of the fact that, on average, their homes reside on more land. Additionally, the monthly lot fees for ROC's tend to drop after the ROC has been in operation for 11 or more years, after which average monthly lot fee for ROC's drops to about $242. In contrast, monthly lot fees tend to go up over time in IOC's (Ward, French, & Giraud, 2005).

Albeit there are certainly other costs associated with living in a manufactured housing park aside from monthly lot fees. However, because lot fees are somewhat less on average for ROC's than IOC's, and because these fees tend to go down over time, this suggests that there is more stability with regard to housing costs in ROC's. It is in ROC residents' own interest to keep costs down, as each owns a collective share of the park.

Opportunity to Build Equity

The third hypothesis, that ROC's provide residents more opportunities to build equity than IOC's, was tested by comparing homes' assessed value, as well as the sale price of homes sold recently in both IOC's and ROC's. The assessed values of individual homes were pulled directly from tax cards, while data on recent sales of homes in the study-sample parks were derived from the local Multiple Listing Service (MLS).

As Table 2 shows, homes in ROC's sold for $4566 more, on average, than homes in IOC's between 1999 and 2005. Part of this price differential might be attributed to the fact that the homes tend to be slightly larger. However, just looking at sales from 2004 to 2005, the price differential increases to $7234. Paul Bradley, with 'The Loan Fund' believes that this is due to the fact that potential homeowners, lending institutions, and other supporting organizations are finally realizing the financial advantages that ROC's pose to homeowners by virtue of the fact that each owns a share of their park.

Table 2
Data from ROC and IOC Sales

	Sales since 1999		**Sales 9/22/04 - 9/22/05**	
	ROC	**Investor**	**ROC**	**Investor**
price	45,884	41,318	53,077	45,843
living area	1035	953	1017.8	936.9
age of home	22.4	22.8	17.6	23
assessed value	38,803	35,565	40,021	36,882
days on market	68	72	60	83
price per sqft	42.4	41.9	55.1	48.6
assessed value per sqft	36.9	36.8	38.7	38.5

Adapted from S. Ward, S., French, C., & Giraud K. (2005). The effect of cooperative ownership on appreciation of manufactured housing. *Cooperative Housing Journal.* 2005/2006 Annual Issue, p. 22.

The fact that homes in ROC's spend, on average, 23 fewer days on the market than homes in IOC's suggests that they may be more desirable to potential homeowners. And, the more demand that there is for a particular housing sector, the more likely it is that it will increase in value. This provides existing and future homeowners in ROC's with enhanced opportunity to build equity, something which lending institutions in New Hampshire are beginning to recognize in their lending habits.

Sense of Ownership and Control

The data collected via mail survey and interviews were coded and analyzed using NVIVO 2.0, a qualitative analysis software package. The results of the analysis suggest that ROC's pose a number of social and economic advantages to park residents as compared to IOC's. Foremost of these benefits is the increased sense of ownership and control over their homes and their communities that is perceived by ROC residents.

A primary reason why residents pursued the formation of a co-op was to gain a greater sense of control over their park and over their lives; they did not want their community to be subject to someone else's decisions. As one co-op Board member described it:

> "I am a part owner of this whole community. I have a say in everything that goes on here whether I'm on the board or not…that is something that you don't normally have in a mobile home park…I own this".

That same individual suggested that this sense of ownership was *not* present when an investor owned the park just a few years prior.
The interview data also suggest that ROC residents are motivated to take care of their homes and their yards because they own them. In fact, when ROC Board members were asked how the physical appearance of their park has changed since transitioning to a cooperative, the majority responded that their park improved. One ROC Board member said:

> "I've heard more from outside people how much nicer the park looks since we've taken over…a lot of the changes are gradual changes…[p]eople that don't come in here often are the ones that notice the difference".

This sentiment was echoed by the town officials that were interviewed, most of which stated that ROC's were better maintained.

Lastly, it is important to note that many of the ROC Board members interviewed in the study have become involved with informal leadership roles as a direct result of their experience on the board. Such roles include coordinating volunteer beautification projects, organizing social events, and heading up a recreation committee. Others have taken on formal leadership roles, including serving on the Parent Teacher Association, Planning Board, Town Council, emergency services coordinator. One Board member became so well known for her success at advocating for her ROC with the state legislature that she was subsequently elected as President of the Manufactured Home Owner Tenants Association of New Hampshire (MOTA).

Conclusion

Based on the analysis of data acquired via personal interviews, tax records, and the Multiple Listings Service, we conclude that resident ownership provides a range of economic and social benefits. Foremost, resident-owned communities provide homeowners with greater access to mortgage financing, whereas homeowners in investor-owned communities are often limited to securing personal property loans or variable-rate loans at a significantly higher interest rate.

A second benefit that resident ownership provides to homeowners is the stability in monthly lot fees compared to monthly lot fees paid by homeowners in investor-owned communities. Not only are the fees lower on average in resident-owned communities, but they also appear to decrease over time. Fees in investor-owned communities generally go up over time.

Resident ownership also appears to have positive implications on home values. On average, manufactured homes in resident-owned communities are valued 10% higher than homes in investor-owned communities, and perhaps even more so over the last couple of years, as new lenders have made mortgage loans available to ROC's.

Finally, ROC's pose a number of social advantages, such as residents' increased sense of ownership and control over their homes and their communities that is manifest in how they take care of their homes and yards.
These findings suggest that resident ownership could be an important sectoral strategy to help low and middle-income families attain social and economic well-being. That is not to say that resident ownership goes without its challenges, such as negative stereotypes that many have of the manufactured housing sector, and the financial and organizational challenges that self-management poses. However, if New Hampshire is any indication, then perhaps the resident ownership model could help manufactured homeowners around the country achieve social and economic well-being.

Implications for Extension

The resident ownership model poses a number of advantages to homeowners. In spite, the model has not taken off across the country because most states provide little in the way of technical, financial, and managerial support to help manufactured homeowners form ROC's and manage them once they are established. The fact is, there are a host of financial, managerial, legal, and infrastructural challenges involved with starting and managing ROC's. Overcoming many of these challenges requires material resources. Perhaps equally important, overcoming these challenges requires a high level of technical and organizational skills that are not likely to be maintained by the residents of manufactured home communities without some external support.

Given that Cooperative Extension specializes in providing individuals, organizations, and communities with educational outreach, perhaps there is an opportunity for Extension to provide ongoing training to ROC Board members (or potential Board members) to build their capacity to address complex issues pertaining to ROC management and thereby increase their likelihood of achieving success through the resident-ownership model.

Based on interviews with 20 ROC Board members, the most crucial skill-building gaps appear to be in the following areas:

- Organizational management (e.g. running meetings, board decision-making, sharing responsibilities).
- Financial management (e.g. billing, business contracts, accounting).
- Maintenance of infrastructure (e.g. contracting, sewer/water maintenance, landscaping).
- Conflict resolution (e.g. interpersonal relations, addressing park violations, resolving disputes with park neighbors and municipalities).

Cooperative Extension already provides training in the above-mentioned areas. Thus, developing training specific to ROC Boards might simply be a matter of tailoring existing curriculum to address issues that ROC's face. Through the provision of training, Cooperative Extension could help advance the ROC model nationally.
With the exception of one or two states, including Minnesota, Cooperative Extension services have not yet worked extensively with ROC's. Perhaps the concept of resident ownership is so new that Extension has not had time to focus its energies on this important affordable housing sector. Or, perhaps collaborative partnerships have not yet been established with organizations, agencies, and institutions whose expertise is needed to ensure the success of ROC's.

But one thing is certain, if housing costs continue to rise faster than incomes, the housing crisis will only worsen. The resident ownership model provides Extension with an opportunity to get involved at the ground floor in helping low to moderate-income communities to build wealth through home ownership.

References

Apgar, W., Calder, A., Collins, M. & Duda, M. (2002). An examination of manufactured housing as a community- and asset-building strategy. *Report to the Ford Foundation by the Neighborhood Reinvestment Corporation.* Cambridge, Massachusetts: Joint Center for Housing Studies, Harvard University.

Bradley, P. (2002). No longer a secret: The manufactured housing sector can create long-term value for homeowners. *Rural Voices.* 8(2), 16-20.

Dillman, D. (1978). *Mail and telephone surveys: The total design method.* New York: John Wiley and Sons.

Genz, R. (2001). Why advocates need to rethink manufactured housing. *Housing Policy Debate*. 12(2), 393-414. Available at: http://www.fanniemaefoundation.org/programs/hpd/pdf/hpd_1202_genz.pdf

Haurin, D., Dietz, R., & Weinberg B. (2003). The impact of home ownership rates: A review of theoretical and empirical literature. *Journal of Housing Research.* 13(2), 119-151. Available at: http://www.fanniemaefoundation.org/programs/jhr/pdf/jhr_1302_haurin.pdf

Manufactured Housing Institute. (2005). How is a manufactured home different from a site-built home? Isn't a 'manufactured home' just a fancy name for a mobile home? *Frequently Asked Questions.* Available at: http://www.manufacturedhousing.org/lib/showtemp_detail.asp?id=231&cat=1#4

Ward, S., French, C., & Giraud K. (2005). The effect of cooperative ownership on appreciation of manufactured housing. *Cooperative Housing Journal.* 2005/06 Combined Issue, 16-24.

US Census Bureau (2005). *Housing vacancy survey.* Washington, D.C.

AN EMPIRICAL MODEL OF PERCEIVED AMBIGUOUS MORTALITY RISKS FOR SELECTED UNITED STATES ARSENIC HOT SPOTS*

To N. Nguyen
Dept. of Agricultural Economics
Texas A&M University

Paul M. Jakus
Dept. of Economics
Utah State University

Mary Riddel
Dept. of Economics
University of Nevada, Las Vegas

W. Douglass Shaw
Dept. of Agricultural Economics
Texas A&M University

Abstract An individual's behavior is likely to be better explained by subjective risks than by experts' assessment of risks, but problems may arise when subjective risks are imprecise or individuals are uncertain about the risks. This is exacerbated when the "experts" themselves also exhibit imprecision in their estimates. Such is the case with mortality risks and arsenic at lower levels in drinking water. In this manuscript we estimate a formal model for a perceived risk distribution that allows for ambiguity. The model is estimated using data collected from a survey given to a sample of people living in arsenic prone areas in the United States.

Keywords: Ambiguity, Risk distribution, Arsenic. JEL Classification: D8, I12

* Shaw is contact author (wdshaw@tamu.edu). The authors acknowledge funding for data collection and researcher support from the U.S. Environmental Protection Agency, STAR grant program (#RJ8RJ7). They wish to thank Trudy Cameron, J.R. DeShazo, Paan Jindapon, Laura Schauer, V. Kerry Smith, Kati Stoddard, and Richard Woodward for comments and help on the manuscript, the survey methodology, and data collection. The usual caveats apply. Authors after the first (Nguyen) contributed equally.

An Empirical Model of Perceived Ambiguous Mortality Risks for Selected United States Arsenic Hot Spots

Introduction

In this paper we report on estimates from a model of perceived mortality risks that are associated with ingesting relatively low levels of arsenic in drinking water. The sample used in the empirical analysis represents the population living in four selected arsenic hot spots in the United States. Unlike arsenic levels consistent with deliberate poisoning in murder mystery novel settings (e.g. the well-known film *Arsenic and Old Lace*), or as found in other countries outside the U.S. [e.g. Bangladesh – see Opar et al. (2004) and Madajewicz et al. (2006)], where arsenic levels might be extremely high (100 to 500 parts per billion (ppb) or even higher concentrations), our hot spot areas contain arsenic levels in water supplies that are mostly well below 100 ppb. The mortality risks (mainly associated with lung and bladder cancer) at these relatively low levels are similar to environmental and health risks where there is likely to be a good deal of uncertainty. The fact that uncertainty might be prevalent in certain contexts was noted at least as early as 1921 by Frank Knight, and was first deemed "ambiguity" by Daniel Ellsberg (1961)[12] Examining choice behavior in the presence of risk has been an active area of research for economists for many decades. Much of the empirical research has relied solely upon a known probability of a given outcome, with little regard for scientific uncertainties

[12] Uncertainty about risks is similar to, but not the same as uncertainty about an expressed maximum willingness to pay, which has been the topic of many non-market valuation studies (e.g. Alberini, Boyle and Welsh 2003). Other scholars have investigated uncertainty, mostly in the context of financial risks (e.g. see the review by Camerer and Weber 1992)

regarding the "point estimates" of the probabilities. For example, one source of scientific uncertainty regarding mortality risks comes from models that are based upon animal tests rather than human tests; another stems from the need to extrapolate model results beyond the range of the observed data. Alternatively, uncertainty about mortality risks may originate with the person needing to make a risky decision, where individuating factors such as exposure levels to toxins, current health status, or activities such as smoking can affect an individual's actual and perceived risk, introducing measurement error for the former, and uncertainty about the latter.

While economists and psychologists have theoretically explored uncertainties about risks in recent years, researchers have been hampered by a lack of empirical models that allow for uncertain risks, i.e, few have estimated statistical models that formally allow for ambiguity in subjective risks. Subjects who are unable to express certainty about risks are simply empirically unusable.

Rather than discard many observations as unusable, we formally introduce a measure of ambiguity which differs across individuals because of key factors. To do this we estimate a perceived risk *distribution* rather than simply providing a point estimate of the mean or median risk.[13] We introduce a measure of ambiguity for arsenic mortality risks that accompany a person's drinking water supply, allowing each person to differ in this

[13] This is in contrast to Savage's subjective expected utility model where it might be assumed that an individual is rational and makes a conscious or unconscious calculation of expected risk, providing a single point estimate (see discussion in Shaw and Woodward 2008). The typicalassumption is that the individual is most comfortable thinking about the mean. If risks were continuously and symmetrically distributed from zero to one hundred percent, then the median and mean risk would coincide with fifty percent. Some psychologists have suggested that a respondent reporting fifty-fifty chance for a binomial outcome actually means that she or he does not know what will happen (Fischoff and Bruine de Bruin, 1999), i.e., that they are completely uncertain.

measure. In this way, we can hypothesize why different people may experience different degrees of uncertainty, similar in spirit to many other researchers who have explored why people have different assessments of risk (e.g. Dosman et al. 2001).

In our empirical application, exposure to arsenic increases the baseline risk of dying from lung and bladder cancer and can cause other health problems in both adults and children. The exposure-response relationships are imprecise, but scientists do believe that health risks resulting from exposure to arsenic are greater for people who smoke or are employed in certain occupations. This suggests that the median and variance of the perceived risk distribution for these people will differ from that of nonsmokers or those not employed in risky occupations. Survey techniques are used to elicit perceived risk and uncertainty and collect general demographic data from individuals exposed to arsenic-contaminated drinking water in the United States; these data are then used to estimate a perceived risk distribution for arsenic exposure.

This manuscript makes three contributions to the literature. First, it provides another, though still rare, empirical application of new approaches in modeling health or environmental risks that formally account for ambiguity in risk perceptions. Second, our model simultaneously estimate the mean and variance (ambiguity) of the risk distribution, formally allowing for heterogeneity across observations in both distributional moments. This is in contrast to the relatively ad hoc methods used to capture ambiguity in past studies. Third, we provide an empirical study that may guide arsenic risk-reduction policy and programs in the United States. The current paper is one of the first studies we know of that focuses on U.S. arsenic exposure from more of a socio-economic, rather than a toxicologic or epidemiologic perspective.

The organization of the remainder of the manuscript is as follows. The next section briefly reviews some relevant literature on arsenic-related health risks and on the topic of uncertain risks, or ambiguity. Section 2 summarizes the survey methodology and the approach used to elicit perceived risks from respondents while Section 3 presents an empirical model for the perceived risk distribution. Empirical results are discussed in Section 4, with conclusions and an outline for future research appearing in Section 5.

1. Background and a Brief Literature Review

1.1 Arsenic and health risk

Arsenic has long been known to be an acute toxin, especially in very high doses. At much lower concentrations (in the range of 50 to 100 ppb), scientists have documented that long-term consumption of arsenic-contaminated water can cause skin damage, problems with circulatory systems, and most seriously, it can increase the risk of contracting lung or bladder cancer. Drinking water (especially ground water) can become contaminated with arsenic at these levels from a variety of sources, including naturally-occurring geologic deposits and from agricultural and industrial practices. For many years the U.S. regulatory standard for arsenic in drinking water was 50 parts per billion (ppb), but in January 2006 the US Environmental Protection Agency (EPA) tightened the standard to 10 ppb. The new regulatory standard provided additional protection to about 13 million Americans in areas with naturally-occurring arsenic in their water supplies (US EPA, 2006).

The tightened arsenic standard was not without controversy. Though scientists agree that exposure to arsenic can damage human health, the exact dose-mortality relationship remains uncertain (*e.g.*, Burnett and Hahn 2001; Wilson 2001). Estimates of the increased health risks that accompany a 50 ppb standard relative to a 10 ppb standard vary (see further discussion and references in Shaw *et al.*, 2006). The dose-response relationship is especially uncertain at low levels (*i.e.*, <10 ppb) leading some scientists to argue that even the 10 ppb threshold is not low enough to ensure safety, although the 2001 report to the EPA by the National Research Council (NRC) showed little evidence of health risks at very low doses. Some [*e.g.*, Burnett and Hahn (2001)] have raised concerns about the data and the methodology used by the EPA to estimate the risks of low-level exposure, and doubt is often cast on inferences for human effects based on animal and epidemiological studies (*e.g.*, Wilson, 2001). In addition, some critics believe that the dose-response relationship from arsenic should be nonlinear and that the actual risk from low level arsenic concentrations is much less than predicted using EPA's linear dose-response model.

The dose-response relationship is further complicated by other factors such as consumption and exposure thresholds, endogenous confounding influences, the latency period for contracting diseases, and the recovery period once exposure to arsenic ceases. Some biologists and toxicologists insist there is an arsenic concentration below which no human health effects are caused, a threshold that is not reflected in current modeling. In addition, endogenous risk-related choices by people, such as cigarette smoking, can further confound the dose-response effects of arsenic exposure. Given similar levels of exposure, the NRC (2001) has found that smokers may be at least twice the mortality risk of non-smokers. The obvious connection here is lung cancer, though smokers are also often at higher risks than non-smokers for many other types of cancer (see Samet 2001). Finally, the mapping of arsenic exposure to health risks is made even more problematic because there is no consensus on the latency period following arsenic exposure, nor the amount of time to recovery after arsenic exposure has ceased. This is important because the latency period can influence values for risk reduction (see Alberini et al. 2006, who consider a delay in risk reductions, as opposed to the delayed onset of disease).

The ongoing scientific debate concerning arsenic and health risks contributes to public confusion about the risks from ingesting arsenic through drinking water. A standard procedure used in estimating people's perceived risk is to inform them using the best available data and information, but with arsenic risks this information may well exacerbate, rather than reduce, confusion about the risks. If scientists hold ambiguous and heterogeneous beliefs about the risks of arsenic, then it is quite possible that the general public will also have ambiguous and heterogeneous perceptions of risks. Thus, economists must address ambiguity and heterogeneity in that ambiguity when modeling consumers' behavior in response to perceived risks.

1.2 Ambiguity and heterogeneity

The pioneering work on risk perceptions by Slovic (1987) and similar work by other psychologists and decision theorists has demonstrated why one should examine perceived risks rather than objective probabilities, especially if one wishes to predict an individual's behavior in the presence of risks. One strand of the risk perception literature suggests that people are simply poorly informed and that "better" risk communication methods will get people to express their subjective risks in accordance with what the scientist or "expert" would dictate. Other scholars remain skeptical regarding drawing conclusions about differences between subjective and so-called expert risks (Rowe and Wright 2001).

While the literature specifically focused on drinking water behavior includes a large number of studies that incorporate measures of perceived risk (e.g. stated point estimates of risks that people provide), these studies often do not involve any ambiguity.[14] The notion of ambiguity goes beyond just incorporating risk perceptions, as several researchers have certainly done (e.g. Poe and Bishop 1999). Daniel Ellsberg (1961) defined ambiguity as the "quality depending on the amount, type, reliability, and 'unanimity' of information." More specifically, Frisch and Baron (1988) define ambiguity as uncertainty about a probability, where uncertainty is created by missing information. Others have stated that this means there is a *second-order* probability distribution, though the "probabilities of probabilities" concept has been controversial (De Finetti 1977; also see discussion in Camerer and Weber 1992). Still others have attempted to explain underlying psychological reasons for exhibiting reactions to ambiguous information (Heath and Tversky 1991; Fox and Tversky 1995).

Since the early 1960's the notion of ambiguity has been extended and analyzed theoretically (*e.g.*, Segal 1987) and in numerous laboratory experiments involving financial gambles and lottery choices under unknown risks. In a very simple construction of an experiment, a subject might be told that there are two or more sources of information about risks, and that these are different. Quite often subjects do not simply average the information to arrive at one expected risk (e.g. Viscusi, Magat and Huber 1991, find that risk perceptions increase as risks are viewed to be more ambiguous). Curley and Yates (1985) found that when constructing lotteries with uncertain

[14] These are reviewed elsewhere (Nguyen, 2008)

probabilities, people exhibited more aversion to ambiguity when the probability interval center increased.

Ambiguity about morbidity or mortality risks has received much less attention than financial risks, despite the fact that that these kinds of risks influence important decisions such as occupational choice, the purchase of insurance, and many everyday activities such as transportation choices. The fact that mortality and morbidity risks vary across people is well known: the Center for Disease Control (CDC) provides one source of "objective" point estimates of mortality risk for the average person in the U.S. population for a large number of causes, but also regularly updates age and gender-specific estimates of mortality risk for several diseases.

People asked to determine their own probability of dying from a particular cause at a particular age can find such exercises difficult because of a confounding emotional response to the prospect of their own death or a lack of information about the objective hazards. So, even when presented with actuarially-sound mortality statistics such as those of the CDC, there is good reason to think that ambiguity may be prevalent in subjective estimates of mortality risks because people with different backgrounds and behaviors, such as smokers, may respond to and assess such risks differently from the general population (*e.g.*, Smith, *et al.* 2001; Viscusi, Magat and Huber 1999).

While many have found that individuals are averse to ambiguity about risks (Viscusi, Magat and Huber 1991; 1999), it is quite possible that in certain settings they have an affinity for it. Cameron (2005) recently formally introduced ambiguity about environmental risks and Riddel and Shaw (2006) similarly introduced ambiguity relating to mortality risks, in somewhat rare empirical behavioral models that do not rely upon laboratory experiments. Cameron (2005) directly elicited the central tendency and variation of global-warming risks by simply asking subjects in her sample (college students) for "…'plus' and 'minus' amounts relative to the expected value…" These values were then interpreted as a four standard deviation range. Riddel and Shaw (2006) allow mean perceived risk to exhibit heterogeneity across respondents using a beta distribution, but then simply use the range of these point estimates as a measure of the ambiguity of respondents. From a statistical point of view, the empirical models are sensible and encouraging in that each allowed for probabilities to be non-linear, allowing a break from the standard linear-in-probabilities expected utility model (EUM). Nevertheless, it would be preferable to estimate the first and second moments of the risk distribution jointly.

Somewhat more formally than either study above, Lillard and Willis (2001), hereafter LW, introduce a probit function approach to model the relationship between the respondents' knowledge of a risky outcome and the shape of the underlying probability density function. Similarly, Hill, Perry and Willis (2005) use Health and Retirement Survey (HRS) data to extend LW's initial work by estimating the determinants of individual-level uncertainty about personal longevity. HPW estimate the shape of the probability density function on the basis of cross-sectional responses to questions that ask only for a point estimate of risk. To do so, they must make rather strong assumptions about what constitutes an exact response or one reflecting subject uncertainty. They conclude that HRS respondents exhibit considerable individual-specific heterogeneity in survival risks and in uncertainty about their true survival probability.

Very recently Riddel (forthcoming, 2008) applies parts of the HPW approach in her test of various specifications for subjective risk distributions, using a data set on perceived mortality risks from nuclear-waste transport. Rather than simply inferring the risk distribution for stated point estimates, she elicits probability ranges using the rungs of a risk ladder and finds that subject's perceptions of the mortality risk from nuclear-waste transport vary with the likelihood of exposure and individual characteristics such as gender. She also finds substantial ambiguity about risks that varies with the amounts and types of information to which subjects have been exposed.

Like Riddel (2008), our approach has the advantage of eliciting ambiguity far more directly than in previous attempts: respondents could provide a point estimate of risk or, if desired, a range within which the point estimated resides. Further, unlike the HRS survey, our study uses standard risk communication techniques (risk ladders and grids) to inform respondents about mortality risks from arsenic and other common causes of mortality. We provide more information about these risk communication devices below.

2. The sample and survey of perceived risks

A complete description of the survey and survey implementation approach is provided in (Nguyen 2008), but here we briefly describe the key features. Our basic sample consists of households living in four communities exposed to arsenic levels in excess of the new EPA standard of 10 ppb at the time of the study (late 2006). While the sample is not representative of the United States as a whole, it was constructed to be representative of the types of people and communities facing risks associated with arsenic concentrations above the EPA standard. Table 1 provides information on the sources of drinking water and arsenic exposure, including the mean and range of contamination, where appropriate, for the four communities. The public water supply systems that provide water to residents of Albuquerque; Fernley, Nevada and Oklahoma City were not in compliance with federal standard for arsenic. The Outagamie County/Appleton Wisconsin region was selected for the study because of arsenic levels in privately owned wells that also exceed the federal standard of 10 ppb. As noted above, private wells are not regulated under the Safe Drinking Water Act, so any knowledge that well owners have about their well quality is obtained on their own, or in conjunction with a state or local health agency.

The survey process involved three key steps. First, random digit dialing was used to recruit participants into the full study. At that time a short telephone survey was given to establish eligibility (households on public supplies that did not pay their water bills were excluded), and collect demographic characteristics. Concerns about environmental risks from atmospheric and water pollutants, and how tap water was used in the household were also questions in this first part of the survey. At the conclusion of the first-round survey a respondent was asked if he or she would be willing to participate in a second-round survey focusing on contamination of drinking water by naturally occurring arsenic. Those willing to participate were sent a four-color and multi-page booklet about the risks of arsenic exposure and how risks can be mitigated, by mail. The mailed booklet asked respondents to consider the risks of arsenic, which were elicited during the follow-up telephone survey (part III).

As with any survey approach, there is possible sample-selection bias. The phone-mail-phone strategy allowed us to explore whether there were differences in the final estimating sample (the sample used to estimate the behavioral model) and the original sample who answered the phone (Part I). We estimated probit models of intended and actual participation in the final study, finding few variables that significantly explain participation. Indeed, we observed few differences in the first round telephone sample and the final participating sample and conclude that selection bias is minimal.[15].

2.1 Explanation and Elicitation of Risk and Uncertainty

The information brochure mailed to each respondent following the initial telephone contact described the sources of arsenic contamination, the effects of long-term exposure, and the new 10ppb EPA standard for arsenic. Following the explanation of the standard, the booklet informed participants of the level of arsenic in the drinking water in their community.[16] Respondents were then provided with detailed information regarding the specific risks of arsenic exposure and the confounding factors that affect an individual's risk such as tap-water consumption, smoking, or exposure to second-hand smoke.[17] Information about baseline risks and risks when arsenic exposure is combined with smoking were communicated with both text and a risk ladder.

Risk ladders have been used for many years as a good device to enhance peoples' understanding of morbidity and mortality risks (see Loomis and DuVair, for example).The risk ladder was designed after extensive focus group work to clarify what problems people had in understanding it. Concerns with early versions of the ladder were that it was too "busy" and cluttered with information, and it was clear that people did not comprehend small numbers (less than 0.01) such as the actual probabilities. That has

[15] Detailed results are provided in another paper and are available on request of the authors (see Cai, et al. 2008).

[16] Those receiving water from a public water system were provided with the mean and range of the arsenic concentration as measured by the local utility. Those on private wells that had not been tested were provided with the range of concentration known to exist in their community.

[17] Other factors included the amount of tap water that a person drinks, the use of an appropriate filtration system, and one's current health status.

been found in other studies (e.g. Hammitt and Graham 1999). Mortality risks are expressed as the number of deaths per 100,000 people in the population**.** It ultimately included fewer “rungs” indicating that baseline lung cancer mortality risk unrelated to arsenic exposure was about 60 per 100,000 people, the risk at 50 ppb was about 1000 per 100,000 people, and the risks of a smoker exposed at 50 ppb was about twice this, at 2000 per 100,000 people.

The final version of the ladder also included rungs associated with other common and not-so-common risks, ranging from death by lightning strike to death from heart disease. While risk ladders are criticized by some as influencing the perception of risk because of scaling “tricks” (e.g. Weinstien 1999; Sandman, Weinstein and Miller 1994), this same type of criticism can be levied at any risk-communication method, and risk ladders do provide a simple means of communicating key information for those who may have little prior knowledge of risks.[18]

Respondents were then asked to think about the mortality risks from arsenic exposure for themselves as well as for other family members, and to express their best estimate of the mortality risk at current exposure levels. Each respondent was asked to put a single mark on the risk ladder if they are certain about the risks; if the respondent could not provide a point estimate of risk, they were asked to place two marks on the ladder, for lowest and highest values of risk. During the second-round telephone interview, respondents were asked which rungs they had marked on the ladder. The survey protocol allowed for respondents to talk about risks with the telephone

[18] Some respondents wanted to think about changes in risks and for them, the risk grid worked better than the risk ladder. Therefore, the final mail brochure included both the risk ladder, and a risk grid depicting the change in risks that they could experience through a risk-reduction program.

interviewer, where interviewers were provided with scripted responses to commonly asked questions.[19]

2.2 Sample and basic statistics

The full sample included 353 households, of which 69% obtain their water from a public water system and 31% from private wells. About 65% of households reported drinking at least some water from the tap but this alone does not indicate arsenic exposure because many do not carefully consider their behavior (see Shaw, Walker and Benson 2005). When asked if tap water was also used to make beverages such as juice or coffee, some 85% responded yes. Survey data indicate that households consume only 24% of their drinking water from bottled water, again suggesting that more than 65% of our households get direct arsenic exposure from tap water. In addition, many households (about 52%) also report that they treat their drinking water.

The risk elicitation process was relatively successful, as compared to some other attempts with which we are familiar. Of the 353 people who completed the final telephone survey, 198 (56%) provided a point estimate of risk, 99 (28%) provided a range of risk, and 56 (16%) could not (or refused) to provide an estimate of risk. Figure 1 shows a histogram indicating the mortality risks as reported by respondents, where the midpoint of the range is used for respondents with ambiguous risks. The distribution is weighted heavily toward relatively low risk levels, with the median of the distribution at about 175 deaths per 100,000 people. This median risk estimate lies between the baseline risk of 60 deaths from bladder and lung cancer and the risk associated with an arsenic

[19] Twenty-seven people initially could not decide how to mark the risks; interaction with the interviewer allowed 22 people to provide either a point or ambiguous estimate of risk. A simple probability model revealed that older people were more likely to not provide a response to the risk elicitation questions.

concentration of 50 ppb (about 1000 deaths), and is consistent with the arsenic exposure levels in the communities, all of which lie below 50 ppb. The histogram also indicates some clustering of responses at higher risk levels. Some studies provide subjects with science-based estimates of risk, only to find that then respondents report back subjective estimates that are thousands of times higher than those based on the best-known science. For example, this was the case for a recent study of mortality risks from nuclear-waste shipping and storage (see Riddel and Shaw 2006), and when asked about this, respondents often said they did not "trust" the government estimates. However, in our case here, it appears that respondents modestly adjust their perceptions in comparison to the science-based risk estimates.

Table 2 presents the basic statistics for the key variables used in estimating the empirical model of risk. Some 57% of sample respondents were *Male* and the average *Age* in the sample was 51 years old. The *Education* survey question is categorized into 7 levels, with no high school attendance as lowest educational level and receipt of an advanced university degree as the highest. The majority of the sample (67%) attended a post-secondary educational institution, with few failing to receive at least a high-school degree. Education was thought to influence risk responses, though the direction of influence is an empirical issue. While cognitive ability may increase with education and lead to a better understanding of small risks, awareness of the complicated nature of the risk issues may also increase an individual's sense of ambiguity about them.

Respondents were asked to rate their own *Health Status* from excellent to poor. Over 64% of the sample rated their current health condition as being very good or excellent. Health status may also influence perceived risks. Again, the direction of influence on one's subjective risk perception is not obvious: people in poor health may feel that they have little to risk from exposure to a toxic substance; in contrast, people in good health may similarly believe they can avoid consequences of exposure. The potential for endogeneity in self-assessed health status has been noted elsewhere (see Moore and Zhu 2000).

Some types of jobs have been shown to increase mortality risks from arsenic exposure because of occupational exposure to other toxins. The survey thus asked the respondents if they were currently employed or had been employed in occupations that scientists believe may increase baseline lung and bladder cancer risks, such as manufacturing paint, textiles, leather, dyes, rubber products or other chemicals, working as a beautician or hairstylist, or working in the printing or aluminum industries. The effects of such occupational exposures were outlined in the mailed information brochure. As shown in Table 2, some 26% of respondents worked or still work in a *Risky Occupation*.

Questions about smoking behavior were taken directly from the HRS survey so the verbal format and presentation of these questions is identical in our survey. About 46% of the sample reported being a *Former or Current Smoker*. Former smokers were defined as anyone who had smoked 100 cigarettes or more in the past, but who did not currently smoke. *Current Smoker*s comprised about 13% of the sample. Again, recall that the information booklet informed respondents that smoking was believed to increase the mortality risks associated with arsenic in drinking water substantially. This was communicated both in text and depicted using the risk ladder. Smoking status, assumed to be exogenous to other explanatory factors in the model, is expected to have a significant effect on an individual's beliefs about their arsenic-related mortality risks, but this is an empirical issue. Smokers often do have different perceptions of smoking-related risks than non-smokers (e.g., see Chapter 4 in Viscusi (2002).

3. Modeling Perceived Risks

3.1 The probit function approach

Any probability distribution with support bounded by zero and one is a candidate for use in modeling perceived risk responses, but we choose to model the subjective risk perceptions and ambiguity using the probit functional form introduced by LW (2001). This is consistent with earlier explorations of the role that the center and range might play in determining ambiguity preferences (see Curley and Yates, 1985). Denote individual *i*'s probabilistic belief about his or her own mortality risk of arsenic present in their drinking water by p_i.

Consider the index function (Hill, Perry, and Willis 2005):

$$I_i = m_i + u_i - \varepsilon_i \quad \text{where: } \varepsilon_i \sim N(0,1), \ \ u_i \sim N(0,\sigma_i^2). \qquad [1]$$

Here, m_i represents all of the information used to form the person's best estimate about the probability. The standard deviation of u_i, σ_i, represents a summary of the information determining a person's ambiguity about the risk, where ambiguity might relate to, for example, lack of information about a risk or uncertainty with respect to how confounding factors affect risk. The random variable ε_i represents unobserved heterogeneity and/or measurement error on the part of the researcher.

The cumulative induced distribution and density functions, $F(p_i)$ and $f(p_i)$, respectively, are derived by Heckman and Willis (1977) based on the index function:

$$p_i = \text{Prob}(I > 0) = \text{Prob}(m_i + u_i - \varepsilon_i > 0) = \Phi(m_i + u_i). \qquad [2]$$

The distribution function is:

$$F(p_i)=\text{Prob}(p_i < p_i^{'}) = \text{Prob}(\Phi(m_i + u_i) < p_i^{'}) = \Phi\left(\frac{\Phi^{-1}(p_i^{'}) - m_i}{\sigma_i}\right) \qquad [3]$$

The median probability of death is $\Phi(m_i)$ where $\Phi(\cdot)$ represent the cumulative distribution function of the normal. For example, if $m_i = -2.326$, the median perceived accident risk is 0.01 implying a death rate of 1,000 per 100,000. Successively smaller (more negative) values of m_i indicate lower perceived risks. For an individual who is certain about the risk $\sigma_i = 0$ and beliefs about p_i degenerate to a point probability equal to $\Phi(m_i)$.

The distribution can take on a variety of shapes including unimodal and bimodal. When $m_i = 0$ and $\sigma = 1$, $F(p_i) \sim \text{uniform}(0,1)$. When $\sigma_i < 1$, the distribution is unimodal with the mode to the left of the median for $m_i < 0$ and to the right of the median for $m_i > 0$. For increasing values of σ_i relative to m_i, the distribution becomes bimodal with modes approaching 0 and 1. HPW (2005) interpret the bimodal case as perfect uncertainty, giving rise to focal responses such as a 50% survival probability, actually the minimum of the distribution.

3.2 Likelihood function

For convenience in notation, the subscript *i* will be suppressed in all that follows below. However, note in fact that it is central to understanding the contribution of this approach to recognize that individual factors can explain both the person's median, and variance of their perceived risk. In this way the variance has individual-specific heterogeneity. The median perceived risk, *m*, and ambiguity as measured by the standard deviation, σ, can be parameterized to derive an estimable model. To do so, allow the median and the standard deviation be given by:

$$m = X\alpha \qquad [4]$$

$$\ln\sigma = Z\beta \qquad [5]$$

where X and Z are variables that influence people's subjective assessment about median and variance of risk respectively. The vectors X and Z need not be identical and may share some elements; α and β reflect weights that the individual put on factors in X and Z. Note that the form in [5] as a semi-log ensures that the standard deviation σ (and, hence, the variance) will be positive.

Substituting [4] and [5] into equation [3] yields the distribution of the risk belief in terms of causal factors for which we have data:

$$F(p) = \Phi\left[\frac{\Phi^{-1}(p) - X\alpha}{\exp(Z\beta)}\right] \qquad [6]$$

with the probability of risk p falling within a range due to ambiguity, $p \in [p_1, p_2]$, specified as:

$$prob(p_1 \leq p \leq p_2) = F(p_2) - F(p_1) \qquad [7]$$

Given the way risks were elicited, an assumption is needed to link an individual's risk responses to $F(p)$. We assume that a reported probability range of $[p_1, p_2]$ implies the person believes that the probability mass lying outside this range approaches zero. This assumption is quite intuitive, stating that the distribution of perceived risk has greatest mass within the stated range. Further, we treat a probability point response as a special case of range response in that the range is bounded by the two midpoints from the rung chosen and its adjacent rungs. The likelihood of point estimate $p = p_o$ can be calculated as a special case of [7] where two ladder rungs are degenerate at a point half way from the marked rung of the ladder and the two next closest rungs. This can be written as:

$$prob(\, p = p_o) = F(p_{ou}) - F(p_{ol}) \qquad [8]$$

where p_{ou} denote the mid-range from p_o to the risk at the next upper rung and p_{ou} denote the mid-range from p_o to the risk at the next lower rung on the risk ladder. Multiplication of [7] or [8] over the appropriate respondents, *i.e.*, those who provide a point estimate versus a range estimate, yields the likelihood function for the entire sample under this first assumption. Ladder rung marks reported from the brochure during the telephone interview were converted to the deaths per 100,000 to correspond with numerical estimates of probability. Maximizing this function will yield the maximum likelihood (ML) estimates of α and β.

4. Estimation results and discussion

Table 3 reports the results for the median and standard deviation of perceived risks from exposure to arsenic via drinking water. The top portion of the table shows the factors affecting the median of the distribution whereas the lower portion shows the factors affecting the variance of the distribution (the ambiguity). While the models appear to be separate in presentation, it is important to remember that all the parameters are estimated using one likelihood function. In this way, estimating of the affect of a variable on the median also takes into account the affect of a variable that explains the variation in the variance.

Among the variables expected to affect the median of the distribution of perceived risk are the arsenic concentration in the community, and a person's smoking status (former, or current - the default category is a person who has never smoked), health status, age, gender, treatment of drinking water at home, and current or past occupation in a risky industry. A positive sign on any of these variables indicates higher perceived risk. The factors believed to influence the standard deviation of perceived risk, or ambiguity, include the range of the arsenic concentration in the community's drinking water and person's smoking status and education. A positive sign on any of these variables also indicates increasing ambiguity about arsenic risks.

For the median portion of the model, the positive and statistically significant coefficient on *Arsenic Concentration*, measured in ppb, implies that people who live in regions with higher arsenic concentrations perceive a greater risk than those who live in regions with lower levels of arsenic, which makes sense if they understand the connection between dose and risk. Figure 2 further examines this: it compares the estimated risk distribution for those consuming 10 ppb arsenic to those consuming 100 ppb, with all other significant model variables calculated at the sample average. The distribution for high concentration (100 ppb arsenic) exhibits dramatic risk skew, pushing the median perceived risk to 1390 deaths per 100,000. Median perceived risk is much lower for the 10 ppb case, at roughly 71 per 100,000. This is in line with the risks that were communicated to the subjects on the risk ladder.

Both measures of smoking behavior, a key confounding factor associated with arsenic risks, are also statistically significant in affecting the median. Respondents who were identified as either a *Former or Current Smoker* had greater perceived risks than those who had never smoked. The difference across current smokers and nonsmokers is striking. Figure 3 compares the risk distributions for current smokers and those who never smoked, with all other significant model variables calculated at the sample average. Both distributions have significant right skew, with a median risk for smokers of 354 deaths per 100,000 nearly three times that of nonsmokers at 120 per 100,000. Since the *Former or Current Smoker* variable enters the distribution through both the median and the variance, smokers have a higher median perceived risk and exhibit more variability in the risk beliefs.[20] The higher variance means that perceived risk distribution for smokers has more weight in the right tail relative to that of nonsmokers. Recall that the risk ladder used in the questionnaire indicates that a smoker's risk of lung and bladder cancer (related to consuming water with 50 ppb of arsenic) are about double that of nonsmokers. The model result suggests that smokers incorporated the information in the ladder and used it to adjust their risk beliefs, perhaps modestly inflating the mortality consequences of smoking and consuming contaminated water.

The negative and statistically significant coefficient of *Former Smoker* in the median portion of the model suggests an interesting aspect of median perceived risk on the part of those who used to smoke but have now stopped. The net effect on median perceived risk of being a former smoker is the sum of the two coefficients for the smoking variables, *Former or Current Smoker* and *Former Smoker*. While having smoked at all in the past raises perceived arsenic risk relative to a current non-smoker, someone who has quit smoking apparently believes that his or her mortality risk is not only below that of current smokers, but also below that of people who have *never* smoked.

It is difficult to simply reconcile these results with the information contained in the booklet, which clearly explained the serious effects of smoking in increasing risks of arsenic. However, former smokers may use the information that they have on cessation from smoking to process and calculation their lung-cancer related arsenic risk. Many

[20] A one-sided hypothesis test is used because the variance is believed by scientists to increase with smoking.

believe that the damaging and long-term effects of smoking disappear after quitting, though this may only be valid with respect to heart disease risks and not with respect to lung and respiratory illness (see Weinstein, 2001). It appears to us that the current smokers in our sample processed the arsenic mortality risk information in accordance with the scientific knowledge, whereas former smokers tended to overly discount the effects of past smoking. While the risk perceptions of former smokers seem somewhat incongruous, the empirical result is in fact consistent with what several other researchers have found.[21]

The positive and marginally significant (P=0.106) sign of *Health Status* in the median shows that a person in a poorer health (self-assessed *Health Status* = 5) perceives higher risk from arsenic exposure than those who are in better health (for example, *Health Status* = 1). The demographic variable *Age* is also marginally significant (P=0.114). The negative sign of *Age* shows that older respondents perceive less risk from arsenic exposure than a younger person. Other variables do not significantly contribute to explaining the median risk.

The factors other than smoking status that also potentially affect ambiguity are shown in the bottom portion of Table 3. The statistically significant intercept term implies that the standard deviation (and, hence, variance) is not equal to zero, and that respondents do possess some degree of ambiguity regarding the risks of arsenic contamination. Among the remaining variables, the *Range of Concentration* (measured in ppb) and its square are very weakly significant, with P-values of 0.29 and 0.21,

---

[21] After experiencing a health shock, for example, Smith *et al.* (2001) found that smokers, former smokers and nonsmokers update longevity expectations (mortality risks) differently.

respectively. The positive sign of the linear term and the negative sign of the squared term provide weak evidence that ambiguity follows a function that is quadratic in the range of possible exposure to arsenic contamination.

Though few variables in Table 3 are shown to significantly explain the variation in the variance, it is important to remember that effects of the variables on both the median and variance are intertwined. Our results on the smoking variable, which enters both portions of the model, may be consistent with Curley and Yates's (1985) early finding about ambiguity: they found that while there was no evidence of ambiguity avoidance in their study with only an increase in the range of the interval within which probabilities might fall, there was indeed evidence of an interaction effect between the interval center and the range of that probability interval.

5. Conclusions, Caveats, and the Need for Further Research

Many, if not all, mortality risks for humans involve risks about which there is very likely some degree of uncertainty. Uncertainty can exist for both the lay person and the expert [begging the question of what an expert is, as Rowe and Wright suggested (2001)] in the context of many sources of mortality risks. We know of few behavioral studies of choice in the presence of risk that use individuals' perceived mortality risk distributions because such an approach has often been considered too computationally complicated. Instead, most risk-related studies in economics tend to rely upon a simple point estimate of risk for use in behavioral models, or they fall back on one single estimate, based on the science and the law of averages.

The perceived risk model presented in this paper provides a computationally straightforward method for estimating the perceived risk distribution of mortality risks from arsenic in drinking water using standard risk-elicitation methods, and it allows one to explore factors that explain individual differences (heterogeneity) in ambiguity. It can be used in other applications in which uncertain risks are prevalent. Our empirical model is parameterized to include the factors believed to influence both the median and variance of the perceived risk distribution, thus allowing us to estimate a distribution that allows for heterogeneity in perceived risk across each person in the sample.

The empirical findings here are closely aligned with several *a priori* expectations. The model revealed that perceived risk is positively associated with exposure (arsenic contamination) levels and also related individuating factors that should affect perceived risk. In particular, respondents' smoking habits are among the strongest influences that affect the median of an individual's perceived risk distribution: all else equal, those who currently smoke perceive higher risks than non-smokers or those who have quit smoking. This is quite different than examining the smoker's sense of his or her own risks related directly to smoking, but as both arsenic and smoking cause lung cancer, it is consistent with literature that finds that smokers may perceive high risks (e.g. Viscusi, 1990; 2002).[22]

We also found that an individual's current health status and age are marginally significant factors influencing the estimate of median perceived risk and explore influences on the variance of the perceived risk distribution. Contrary to the frequently employed empirical assumption of a zero variance (as in models strictly derived using the traditional expected-utility framework that depends solely on risk), we find evidence that the

[22] To be fair, we note again here that former smokers also believe their risks to be lower than even people who have never smoked, consistent with concerns raised by Slovic (2001).

variance of the perceived risk distribution is non-zero. We find that those who currently smoke, or who have smoked in the past, show greater ambiguity about the mortality risks of exposure to arsenic. The link between the variance of the risk distribution and exposure levels was weak, but consistent with expectations.

The approach used in this paper has important policy implications. First, it is clear that policy-makers need to better address the problem of communicating risks that are scientifically ambiguous. We find that people resond to risk information differently, supporting the notion that risk communication be specifically tailored to certain types of people rather than a one-size fits all approach. Third, the approach here can be applied to allow researchers to bring a richer model of perceived risk and its distribution directly into behavioral models. In contrast to previous studies that only incorporate perceived risks, or to more recent papers that incorporate ambiguity, but in somewhat of an ad hoc manner, the more formal approach used in this paper allows researchers to specify a more general model of perceived risk.

We have not closely examined the issue of latency in the disease as others have (e.g. Alberini et al. 2006), though our respondents were informed that scientists believe the cancers would arise only after prolonged exposure. An important caveat is that there is here again, some ambiguity about the length of time before cancers would occur, and it is possible that this also contributes to the variance in individual’s risk distributions. One of the next steps might be to allow more even more general heterogeneity across people in samples used for estimation, introduced into both the central tendency of risk and its variance. After this, a logical extension of the research is to connect this type of non-expected utility model to a measure of non-market valuation, which has been considered theoretically before (see Smith, 1992; Jindapon and Shaw, forthcoming 2008), but needs to be addressed much more empirically.

References

Alberini, Anna; K. Boyle; M. Welsh. (2003). "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management* **45** (1/January): 40-62.

______________; M. Cropper; A. Krupnick; N.B. Simon. (2006). "Willingness to Pay for Mortality Risk Reduction: Does Latency Matter? *Journal of Risk and Uncertainty* **32**: 231-45.

Bassett, William F. and Robin L. Lumsdaine. (2001) "Probability Limits: Are Subjective Assessments Adequately Accurate?" *Journal of Human Resources* **XXXVI** (2): 327-63.

Burnett, Jason K. and Robert W. Hahn (2001). "A costly benefit." *Regulation*, *24*(3), 44-49.

Cai, Yongxia; W. Douglass Shaw; Ximing Wu (2008). "Altruistic Averting Behavior in Response to Mortality Risks of Arsenic in Drinking Water." Discussion paper, Texas A&M University.

Camerer, Colin & Martin Weber (1992). "Recent developments in modeling preferences: uncertainty and ambiguity." *Journal of Risk and Uncertainty*, ***5***, 325-370

Cameron, Trudy A. (2005). "Individual option prices for climate change mitigation." *Journal of Public Economics*, ***89*** (2-3/ February), 283-301.

Curley, Shawn P. and J. Frank Yates. (1985). "The Center and Range of the Probability Interval as Factors Affecting Ambiguity Preferences." *Organizational Behavior and Human Decision Processes* **36**: 273-287.

De Finetti, B. (1977). "Probabilities of Probabilities: a Real Problem or a Misunderstanding?" In *New Developments in the Applications of Bayesian Methods*, A. Aykac and C. Brumat (eds.). New York: North Holland.

Dosman, Donna M.; W.L. Adamowicz; S.E. Hurdley. 2001. "Socioeconomic Determinants of Health-and-Food Safety Related Risk Perceptions." *Risk Analysis* **21** (2): 307-17.

Ellsberg, Daniel (1961) "Risk, ambiguity, and the Savage axioms." *Quarterly Journal of Economics*, **75**, 643-669.

Fischoff, B. and W. Bruine de Bruin. 1999. "Fifty-fifty = 50%." *J. of Behavioral Decision Making* **12**: 149-63.

Fox, Craig R. and Amos Tversky (1995). "Ambiguity aversion and comparative ignorance." *Quarterly Journal of Economics*, ***110***, 585-603.

Frisch, Debora and Jonathan Baron (1988). "Ambiguity and rationality." *Journal of Behavioral Decision Making*, ***1***, 149-157.

Hammitt, James and J.D. Graham (1999). "Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?" *Journal of Risk and Uncertainty* **8**: 33-62.

Heath, Chip and Amos Tversky (1991). "Preferences and belief: ambiguity and competence in choice under uncertainty." *Journal of Risk and Uncertainty*, ***4***, 5-28.

Heckman, James J. & Robert J. Willis (1977). "A beta-logistic model for the analysis of sequential labor force participation by married woman." *Journal of Political Economy*, ***85*** (1), 27-58.

Hill, Daniel, Michael Perry, and Robert J. Willis (2005). "Estimating Knightian uncertainty from survival probability questions on the HRS." Manuscript, University of Michigan.

Jindapon, Paan and W. Douglass Shaw. 2008. "Ex Ante Welfare Measures without Expected Utility." In Press, *Economics Letters*.

Knight, Frank. (1921). Risk, Uncertainty and Profit. Boston: Houghton, Mifflin Co.

Lillard, Lee A. and Robert J. Willis (2001). "Cognition and wealth: The importance of probabilistic thinking." Michigan Retirement Research Center Working Paper WP2001-007.

Loomis, John B. and Pierre H. duVair. (1993). "Evaluating the effect of alternative risk communication devices on willingness to pay: Results from a dichotomous choice contingent valuation experiment." *Land Economics*, ***69***(3), 287-298.

Madajewicz, Malgosia; A. Pfaff; A. van Green; J. Graziano; I. Hussein; H. Momotaj; R. Sylvi; and H. Ashan. 2006. "Can Information Alone Change Behavior? Arsenic contamination of groundwater in Bangladesh." Discussion paper, Earth Institute, Columbia University.

Moore, Michael J. and Carolyn W. Zhu. (2000). "Passive Smoking and Health Care: Health Perceptions Myth vs. Health Care Reality." *Journal of Risk and Uncertainty* **21** (2/3) 283-310.

National Research Council. (2001). "Arsenic in Drinking Water, 2001 Update." Washington, D.C. 310 pp.

Nguyen, To N. (2008). "An Empirical Model of Perceived Risks for Arsenic Concentrations." Unpublished PhD dissertation chapter, Texas A&M University.

Opar, A.; A. Pfaff; A.A. Seddique; K.M. Ahmed; J.H. Graziano; and A. van Geen. 2004. "Responses of 6500 Households to Arsenic Mitigation in Araihazar, Bangladesh." Discussion paper, Lamont-Doherty Earth Observatory, Columbia University

Poe, Gregory L. and Richard C. Bishop. 1999. "Valuing the Incremental Benefits of Groundwater Protection when Exposure Levels are Known." *Environmental and Resource Economics* **13** (3): 341-67.

Riddel, Mary (2008). "Risk perception, ambiguity, and nuclear-waste transport." Forthcoming, *Southern Economic Journal*.

Riddel, Mary and W. Douglass Shaw (2006). "A theoretically-consistent empirical model of generalized expected utility utility: An application to nuclear waste transport." *Journal of Risk and Uncertainty*, ***32*** (2), 131-150

Rowe, Gene and George Wright (2001). "Differences in Expert and Lay Judgments of Risk: Myth or Reality?" *Risk Analysis* **21** (2), 341-356.

Samet, Jonathan. 2001. "The Risks of Active and Passive Smoking." Chapter 1 in *Smoking: Risk, Perception and Policy*, Paul Slovic (ed.). London: Sage Publications.

Sandman, P.M.; N.D. Weinstein; P. Miller, 1994. "High Risk or Low: How Location on a 'Risk Ladder' Affects Perceived Risk." *Risk Analysis* **14**: 35-45.

Segal, Uzi (1987). "The Ellsberg paradox and risk aversion: an anticipated utility approach." *International Economics Review*, ***28***, 175-202.

Shaw, W.Douglass, Paul Jakus, Mary Riddel, Mark Walker, and Paan Jindapon. (2006). "Incorporating perceived mortality risks from arsenic into models of drinking water behavior and valuation of arsenic risk reductions: preliminary results." Presented at the *U.S. EPA Workshop on Mortality Risks*, Washington D.C., April 10-12, 2006.

Shaw, W. Douglass, Mark Walker, and Marnee Benson. (2005**).** "Treating and Drinking Well Water in the Presence of Health Risks from Arsenic Contamination: Results from a U.S. Hot Spot" *Risk Analysis*, **25** (6): 1531-1543.

Shaw, W. Douglass and Richard T. Woodward (2008). "On Why Environmental and Resource Economists Should Care about Non-Expected Utility Models" In press, *Resource and Energy Economics*.

Slovic, Paul (1987). "Perceptions of risk." *Science*, 236(17 April), 280-285.

____________ (2001). "Cigarette Smokers: Rational Actors or Rational Fools?" Chapter 6 in *Smoking: Risk, Perception and Policy*, Paul Slovic (ed.). London: Sage Publications.

Smith, V.K. 1992. Environmental Risk Perception and Valuation: Conventional Versus Prospective Reference Theory. Chapter 2 in The Social Response to Environmental Risk: Policy Formulation in an Age of Uncertainty, D.W. Bromley and K. Segerson (eds.). Boston: Kluwer Academic Press. 217 p.

Smith, V. Kerry, Donald H. Taylor Jr., Frank A. Sloan, F. Reed Johnson, William H. Desvousges (2001). "Do Smokers Respond to Health Shocks?" *Review of Economics and Statistics*, 83 (4/November): 675-87.

Tobin, James (1959). "On the Predictive Value of Consumer Intentions and Attitudes." *Review of Economics and Statistics* **41** (1): 147-55.

U.S. Environmental Protection Agency (US EPA) (2006). Arsenic in drinking water. Retrieved Sep 01, 2007, from http://www.epa.gov/safewater/arsenic/

Viscusi, W. Kip. (1990). . "Do Smokers Underestimate Risks?" *Journal of Political Economy* **98**, 6: 1253-1269.

_______________ (2002). *Smoke-filled Rooms: a Post-Mortem on the Tobacco Deal*. Chicago: University of Chicago Press.

Viscusi, W.Kip, Wesley Magat, and Joel Huber. (1991). Communication of Ambiguous Risk Information." *Theory and Decision* **31**: 159-73.

____. (1999). "Smoking status and public responses to ambiguous scientific risk evidence." *Southern Economic Journal* **66** (2): 250-70.

Weinstein, Neil D. (1999). "Accuracy of Smokers' Risk Perceptions." *Nicotine & Tobacco Research* **1** (Supplement): S123-S130.

________________. (2001). "Smokers' Recognition of their Vulnerability to Harm." Chapter 5 in *Smoking: Risk, Perception and Policy*, Paul Slovic (ed.). London: Sage Publications.

Wilson, Richard (2001). "Underestimating arsenic' risk." *Regulation*, *24*(3): 50-53.

Figure 1: Distribution of risk responses (point and mid-range, deaths per 100,000)

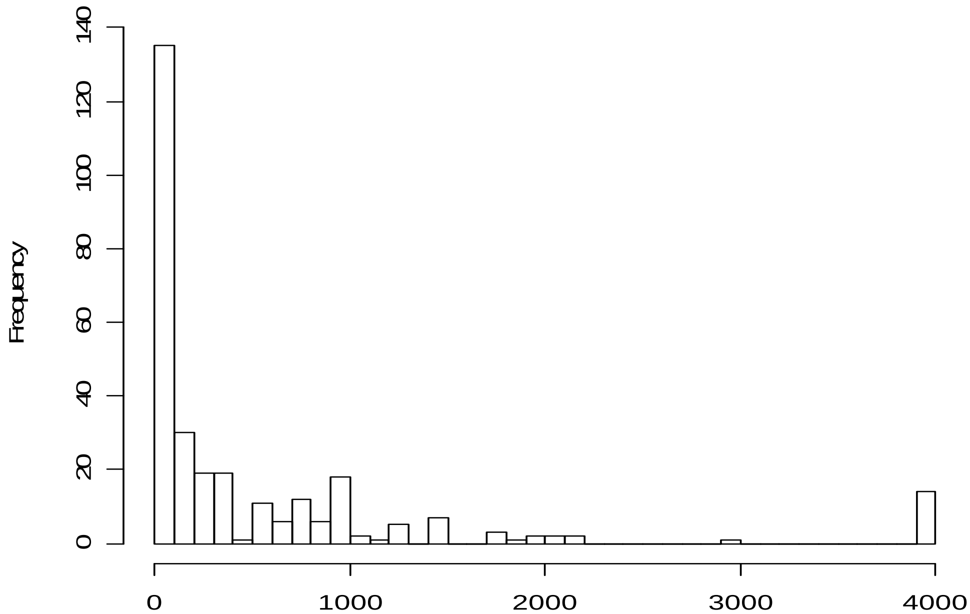


Figure 2. Estimated probability density function of perceived risk for areas with arsenic concentrations of 10 ppb and 100 ppb.

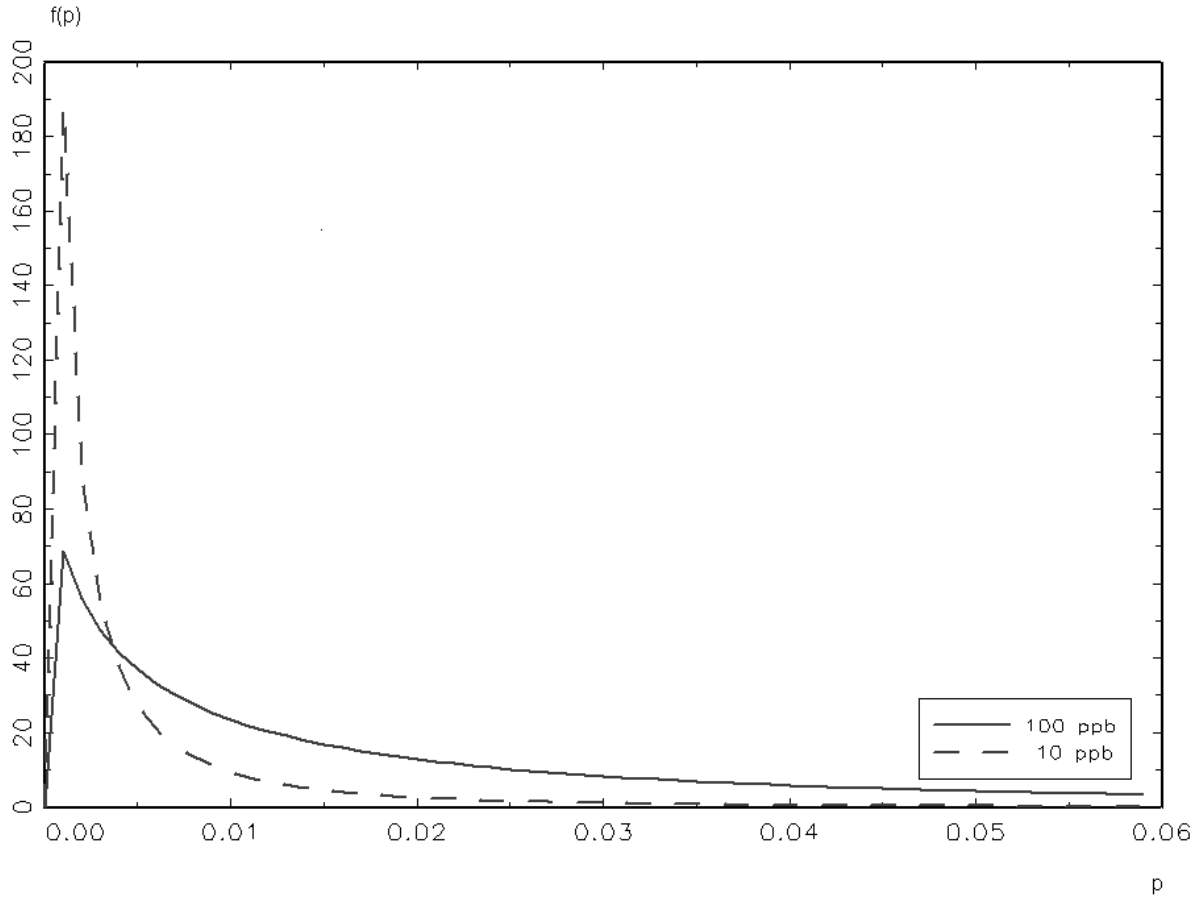

f(p)
0
20
40
60
80
100
120
140
160
180
200
0.00
0.01
0.02
0.03
0.04
0.05
0.06
p
100 ppb
10 ppb

Figure 3. Estimated probability density function of perceived risk for current smokers and those who have never smoked

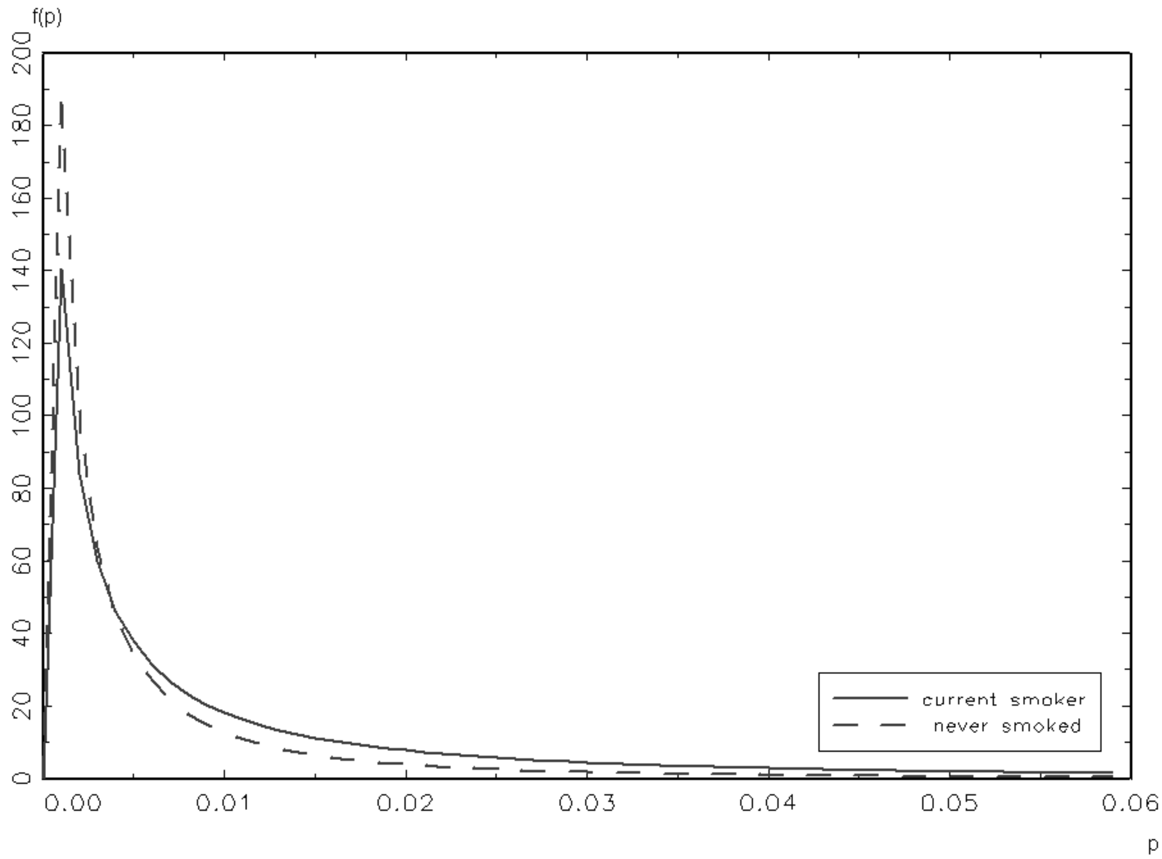


Table 1: *Profile of Water Sources, Arsenic Concentrations*

Area	Water Source	Mean Arsenic Concentration (ppb)	Range of Concentration
Albuquerque, NM (n=54)	Public	25	20 – 30
Fernley, NV (n=108)	Public	40	No range
Oklahoma City, OK (n=80)	Public	17.5	14 – 21
Outagamie County, WI (n=55)	Private, tested	3.84	No range
Outagamie County, WI (n=43)	Private, not tested	—	5 – 105
Appleton, WI (n=5)	Private, tested	6.9	No range
Appleton, WI (n=8)	Private, not tested	—	5 – 105

Table 2: *Basic statistics of key variables*

Variable	Value	Percent or Mean
Male	1 = yes, 0 otherwise	56.7%
Age	Years	51
Education (n=342)		
No High School	1	4.4%
Some high School	2	3.2%
High School Graduate	3	24.9%
Some College	4	19.3%
Two-year Degree	5	17.8%
Four-year Degree	6	19.0%
Advanced Degree	7	14.3%
Health Status (n=353)		
Excellent	1	27.5%
Very Good	2	36.8%
Good	3	26.9%
Fair	4	7.6%
Poor	5	1.1%
Risky Occupation (n=353)	1 = yes, 0 otherwise	26.1%
Former or Current Smoker	1 = yes, 0 otherwise	45.6%
Current Smoker	1 = yes, 0 otherwise	12.7%
Treat Drinking Water	1 = yes, 0 otherwise	51.6%

Table 3: *ML estimation of median and variance of perceived risk*

Variable	Coefficient	Standard Error	P-value
	Median		
Intercept	−3.239***	0.164	0.001
Arsenic Concentration	0.011***	0.003	0.001
Former or Current Smoker	0.342**	0.131	0.010
Former Smoker	−0.439***	0.145	0.003
Health Status	0.071	0.044	0.106
Age	−0.004	0.003	0.114
Male	0.010	0.084	0.906
Treat Drinking Water	−0.037	0.080	0.643
Risky Occupation	−0.052	0.097	0.592
	Variance		
Intercept	−0.470*	0.163	0.005
Range of Concentration	0.059	0.056	0.288
Range of Concentration^2	−0.008	0.006	0.211
Education	−0.019	0.030	0.531
Former or Current Smoker	0.151#	0.103	0.072#
Log-likelihood: −655.255			

***, **, and * represent significance at the 0.01, 0.05, and 0.10 levels respectively, for a two-tail test.
represents significance at the 0.10 level for a one-tail test.

Experiments with Invitation Designs to Maximize Web survey Responses Rates

Michael D. Kaplowitz, Frank Lupi, Mick Couper, and Laurie Thorp

ABSTRACT

Web surveys increasingly are being used. However, Web surveys present methodological challenges including lower response rates than other survey methods. This paper reports the results of a large-scale experiment to test design features of e-mail and postcard invitations for maximizing Web survey response rates. A stratified, random sample of 15,648 students, faculty, and staff of a major U.S. public university received either a postcard or e-mail invitation to participate in a campus sustainability survey. Using a full factorial design, the postcard and e-mail invitations were modified to vary such elements as their length (long/short), estimate of effort (about 10/less than 30 minutes), and subject line (authority/topic). The results reveal significant effects of invitation design on response rates. After two waves of invitations, before postcard recipients received a final e-mail invitation, the use of postcard invitations (as opposed to e-mail invitations) reduced faculty and to a lesser extent staff response. The invitation mode did not seem to affect response rates for students. After three waves of invitations, the results show that invitations estimating survey effort as “about 10 minutes” as opposed to “less than 30 minutes” increased student response but did not affect faculty and staff response rates. The text length of invitations revealed longer invitations increased faculty and staff response rates and made no difference for students. Furthermore, invitations using the “authoritative” as opposed to “salient topic” subject line significantly increased response from all groups. The results suggest that some design elements of invitations may have different effects on different subsets of potential respondents and therefore should be tailored accordingly to maximize response.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Fred Poston, Michigan State University (MSU) Vice President of Finance and Operations (VPFO) and Kathryn Lindahl, Assistant VPFO for their support, encouragement, and especially for providing a opportunity to integrate undergraduate, graduate, and faculty research into a campus-wide environmental stewardship effort. We also thank Chelsea McMellen, Lauren Olson, Diana Oppenheim, Aimee Wilson, Sibbir Norman, Felix Yeboah, and David Gebben for their research assistance and hard work. Any and all errors and omissions are those of the authors.

INTRODUCTION

The World Wide Web (also called the Web or the Internet) increasingly is being used as a means of surveying the public, special populations, and targeted groups of potential respondents (Dillman 2007, Couper 2000). Advantages of using Web surveys include cost savings associated with forgoing the printing and mailing of survey instruments (Cobanoglu et al. 2001) as well as time and cost savings of having survey data returned already in electronic format (Couper et al. 2001). Furthermore, Web surveys provide researchers with the ability to use complex skip-patterns, question designs, and graphics. Of course, Web surveys require computer programming skills, special software, and hardware capabilities not required for typical mail surveys. Investigators have begun to explicate, test, and otherwise address some Web survey design, sampling, and implementation strategies (e.g., Dillman 2007; Couper et al. 2007; Groves 1989; Schafer and Dillman 1998). Likewise, researchers have begun to examine how responses of sample populations of Web users may or may not be indicative of the general public's responses to other survey modes on substantive variables under investigation (Denscombe 2006; Couper 2000; Dillman 2007; Schaefer and Dillman 1998).

For populations that regularly use the Web as well as other populations that might be amenable to online surveys, Web surveys have been found to be a useful means of conducting research (Sills and Song 2002; Couper, 2000; Couper et al., 2004; Denscombe, 2006; Evans & Mathur, 2005; Huang, 2006; Witte et al., 2000). While the generally accepted elements of the "Tailored Design Method" for mail surveys (Dillman 2007) were the product of years of research and intensive study, the best practices for internet surveys are still a work in progress. Web surveys present methodological challenges including lower response rates than other survey methods. Implementation approaches beneficial for mail surveys may not translate directly to response rate benefits for Internet surveys (Couper 2000). This paper reports the results of a large-scale, full-factorial experiment to test design features of invitations to participate in a Web survey in order to help identify invitation design features for maximizing Web survey response rates.

Invitations to Participate/Advance Letters

The literature on invitations to participate in Web surveys builds on previous research regarding advance letters that suggests that advance letters are cost-effective ways of increasing response rates in mail surveys and interviewer-administered surveys (e.g., Dillman et al. 1995; Hembroff et al., 2005; De Leeuw et al., 2007; Dillman 2007). Common strategies used in such advance letters to participate in paper-based and telephone surveys include monetary incentives, pre-notification, personalization, increased salience of subject matter, and other design elements (Dillman 2007, Edwards et al., 2002). However, the investigation of the efficacy and appropriateness of design elements of invitations to participate in a Web survey is relatively new. This paper reports the results of a full-factorial experiment of five design elements of invitations to participate in a Web survey—invitation mode (postcard versus e-mail invitation), subject line (authority versus salient subject), location of URL link/information (top versus bottom), length of the invitation text (short versus long), and the survey time/effort estimate (about 10 minutes versus less than 30 minutes).

Invitation Mode. The mode (e.g., e-mail, SMS (text) message, postcard) used to invite potential respondents to participate in a Web survey may elicit different response rates. A prenotice or advance letter alerts the sample person to an upcoming survey request, but does not contain the questionnaire itself (in mail surveys) or the means to complete the survey (a URL in the case of Web surveys). While there have been several studies of modes of prenotification for Web surveys (e.g., Kaplowitz, Hadlock, and Levine, 2004; Crawford et al., 2004; Harmon, Westin, and Levin, 2005; Porter and Whitcomb, 2007; Bosnjak et al., 2008), there is a paucity of research on the mode of invitation for Web surveys. In one exception, Bosnjak and colleagues (2008) found an SMS invitation to be less effective than an e-mail invitation. In general, mailed invitations are used where e-mail addresses are not available or for mixed-mode surveys (mail with a Web option), while e-mail invitations are believed to be superior when the sample frame contains such information. Our paper examines the effect of a postcard versus an e-mail invitation to participate in a Web survey.

Invitation Length. There is evidence from mail surveys that the length of a questionnaire influences the response rate, with shorter questionnaires yielding a higher response rate (Edwards *et al.*, 2002). Some studies show that the length of Web surveys yield similar results (Cook *et al.*, 2000; Galesic, 2007; Marcus *et al.*, 2007). However, there has been a lack of research pertaining to the length of the invitation to complete a Web survey and response rates. The general advice has been to keep invitations short–people don't want to read too much. To our knowledge, our study is the first to examine the effect of varying the length (and hence content) of an invitation to take a Web survey on response rates.

URL Placement. Like invitation length, the placement of logon information or URL links in Web survey invitations has been more art than science. Some studies have compared URLs with embedded identifiers (automatic or passive authentication) to those requiring manual entry of an ID or password (manual or active authentication) (see Crawford, Couper, and Lamias, 2001; Heerwegh and Loosveldt, 2002, 2003; Joinson, Woodley, and Reips, 2007). However, we are aware of no previous research on the placement of the URL in a Web survey invitation. The general advice regarding the Web survey link or URL has been to keep it short and place the URL "above the fold," based on the belief that people don't want to read much or scroll, and that they make a quick judgment about legitimacy and go straight to the URL. Our paper reports the results of an experiment that tests alternative placement of the Web survey URL in the invitations.

Survey Effort/Length. It has been suggested that if a survey is short, participants will be more likely to take part and complete the survey (Ray & Tabor, 2003; Cook *et al.*, 2000; Edwards *et al.*, 2002; Marcus *et al.*, 2007). One early study (Crawford, Couper, and Lamias, 2001) suggested that more people start shorter surveys, but if the survey takes longer than the announced time, the researchers reported more break offs. Touteaud (2004) found increased response rates when subjects were informed that the survey would take 3-5 minutes versus 10-15 minutes. Marcus *et al.* (2007) reported that a survey respondents were told would take 10-20 minutes resulted in a higher response rate (31%)

than one where they were told it would take 30-60 minutes (19%). Our research builds on this previous work by testing alternative time/effort estimates for a Web survey sent to three segments of university community members.

Subject Line. There has been limited research on the effects of different subject lines in email invitations to surveys, all yielding different results in terms of response rates. Porter and Whitcomb (2005) examined low involvement and high involvement subjects' response to e-mail invitations that varied subject line elements. They reported increased response (decreased click-through) for invitation messages providing the reason for the e-mail invitation as well as invitations identifying the survey sponsor, while e-mail invitations simply requesting assistance appeared to have no response effect. Trouteaud (2004) tested a "plea" subject line versus an "offer" subject line in a survey invitation to subscribers to an online newsletter and reported a 5 percentage response rate advantage for the plea version. Damschroder and colleagues (personal communication) tested two alternative subject lines in an e-mail invitation: "Participate in an important study on health issues" versus "University of Michigan Health Study" and found no difference in the response to the invitation. Use of a high authority figure in the invitation has been suggested to increase survey response rates, so using an authority figure in an invitation's subject line could yield similar results (Gueguen and Jacob, 2002; Joinson and Reips, 2007). It has also been shown that surveys with a high subject matter salience can yield higher response rates (Cook *et al.*, 2000; Edwards *et al.*, 2002; Marcus *et al.*, 2007). We test response rate effects of invitations using a salient subject matter line versus those using an authoritative request in the subject line.

Research Context

The reported research is based on a larger research effort at Michigan State University (MSU) focused on campus sustainability. In 2006, MSU began systematically studying methods to improve campus sustainability efforts. The MSU Vice President of Finance and Operations embraced implementing MSU's Boldness by Design initiatives including strengthening the university's environmental stewardship. One stewardship goal aims to simultaneously reduce the university's environmental footprint and increase the efficiency of its materials and energy use. As part of this campus initiative, MSU asked campus researchers to help develop and design an improved recycling program. One part of this integrated research, teaching, and outreach effort was an assessment of the MSU community's recycling knowledge, perceptions, use, and program preference. This paper is based on the results a campus-wide survey inviting response from roughly one-third of [university] students, faculty, and staff (*N*=15,648).

The Web survey instrument was developed in multiple phases, in an iterative process (Kaplowitz *et al.*, 2004b). A series of key informant interviews were conducted with university administrators, consulting engineers, faculty active in university environmental affairs, as well as student activists. There were also focus group studies with members of the target audiences as well as an iterative survey design and pretesting/revision process. The Web survey seamlessly leads respondents through a series of questions that only pertained to them based upon their previous answers. Respondents were asked about recycling options where they live and work, their recycling knowledge and attitudes, and their preferences regarding alternative recycling program characteristics. Also, respondents were asked about the effectiveness of various communication media, their environmental attitudes, and some basic demographic information. The survey's substantive results concerning communication strategies as well as results of the preference questions are reported elsewhere [citations omitted]. This paper focuses on the full-factorial experiment of the design elements used in the invitations to potential respondents to participate in the survey.

Methods

Survey Sample, Implementation, Overall Response

The sample list for the study was drawn from the university's official lists of faculty, staff and student. All students, faculty, and staff at MSU have a university e-mail account, free access to the Internet, as well as a mailing address on record. Furthermore, MSU students, faculty, and staff are expected to use their university e-mail address to communicate with instructors, administrators, registrar, classmates, etc. The registrar provided random lists using systematic sampling of the email and mailing addresses of 30% of each MSU group—students, faculty, and staff.

An initial invitation was sent to all members of the sample during November 2007 informing them of the study and providing them with a link to the survey as well as logon information. Up to two additional invitations to participate were sent to those who had not yet responded, with about 10 days separating each mailing. Altogether, the Web survey contained several brief sections of questions with completion of the Web survey taking most respondents about 10 minutes.

Experimental Design and Hypotheses

The invitations were designed so that they could accommodate five experimental manipulations—mode, length, URL, effort estimates, and subject lines (See Table 2 and Figures 1 and 2). Within each of the three strata, sample persons were randomly assigned to one of the 32 versions of the invitation, created by crossing each of the five factors in a full factorial design. For cost reasons, one third of the sample was allocated to the postcard condition, while two thirds were assigned to the e-mail condition. All other experimental conditions were assigned to a random half of each group.

[Table 1 about here]

After two postcard invitations, those in the postcard group who had not yet responded were switched to e-mail for a final (third) reminder. Aside from this, each individual received the same experimental design for every contact. The e-mail invitations were virtually exact copies of the postcard invitations with same text and formatting used for both (see Figures 1 and 2) except that recipients of the postcards had to go to a computer to logon to the survey using the unique URL assigned to them.

[Figure 1 about here]

In addition to the invitation mode experiment, we alternated the length of text used in the invitations, creating a long (Figure 2) and a short (Figure 1) text version, with half the sample assigned to each version. The long invitation consisted of around 182 words, where the short invitation consisted of around 80 words. The additional wording was carefully chosen and tested so that it was neutral, neither persuasive nor off-putting. We expected that, all else equal, a shorter invitation would yield a higher response rate.

[Figure 2 about here]

We designed the invitations so that half of the sample received an invitation with the URL or link to the Web survey placed above the main body of text of the survey invitation while the other half received an invitation with the URL placed below the text of the survey invitation. For the postcard condition, the URL information included requiring recipients to enter the URL in order to participate in the study. A unique URL was assigned to each member of the sample list and designed to be easy to enter by hand should the URL not work by simply clicking on it. We expected that placing the URL near the top would be associated with a higher response rate.

The actual length of the Web survey was kept constant at about 10 minutes for all participants. The time/effort estimate manipulation of the experiment delivered to half of the sample an invite noting that the survey would take "about 10 minutes" while the other half of the sample received an invite noting that the survey would take "less than 30 minutes." We expected that the shorter time estimate would be associated with a higher response rate.

The final design experiment was the subject line manipulation. Half of the subjects received their invitation(s) with a subject line indicating that the message was from an authority figure, the Vice President of Finance and Operations (see Table 2 for exact wording). The other half received an invitation with a subject line indicating that their input on a campus environmental/sustainability was needed. Our expectation is that the response rate to Web survey invites is not influenced by the subject line content.

Results

Aggregate Response Rates

The survey was initially sent to 15,648 individuals representing about 30% of each group of the campus population. A total of 3896 individuals participated in the study. As Table 1 illustrates, the AAPOR minimum response rate (RR1) for the study was 24.9% (AAPOR 2006). There were significant differences in response rates across the three campus sub-populations—response rates were highest for staff (42.8%), followed by faculty (38.0%) then students (20.2%) (χ^2(2,n=15648)=630.85, p=0.000).

[Table 2 about here]

Response to Design Elements

Up to three invitations were sent to the 15,648 members of our random sample lists of students, faculty, and staff. The first two waves of invitations delivered the same invitation design to subjects using the same delivery mode. After that, nonrespondents who initially received their invitation via postcard received their third and final invitation via e-mail. That invitation was otherwise identical in format and design to the initial postcards. In order to understand the impacts of the design elements, including use of mixed-mode invitations, we begin by presenting the results of our analysis of the design experiments for the first two complete waves of invitations.

[Insert Table 3 about here]

Table 3 presents the response rates as well as the results of chi-square tests of differences in response rates by invitation design treatment after two contacts. After two waves of contacts, e-mail invitations resulted in significantly higher response from faculty (12% more, $p<.01$) and staff (4% more, $p<.10$), but students did not respond statistically differently to postcard and e-mail survey invitations. Note too from Table 3 that postcard invitations, by themselves, can still attract respondents to the Web survey, despite the inconvenience of hand typing the URL. The length of the survey invitation itself, after two contacts, appears to contradict the 'conventional wisdom' that shorter is better. As Table 3 illustrates, the long version of the invitation resulted in significantly higher response rates from the faculty (5% more, $p<.05$) and staff (4% more, $p<.10$), while the student response rate did not statistically differ for the different invitation lengths after two contacts. The location of the Web survey URL or link appears to make a difference for faculty and student response after two contacts. The faculty response rate was higher (5% more, $p<.05$) as was the student response rate (1% more, $p<.05$) when the URL was at the bottom on the Web survey invitation. There was no apparent difference in staff response based on the URL location in the invitation.

Interestingly, despite strong priors that telling subjects that the survey would take less time, there were no significant differences in the response rate for faculty and staff for the different survey effort treatments. However, student response to the “about 10-minute” effort estimate treatment was significantly higher (2% more, $p<.01$) than the “less than 30-minute treatment” after two contacts. An examination of the completion rate of subjects that started the Web survey by clicking “I Consent” on the survey’s informed consent page revealed that less than one percent of respondents who consented to participate in the survey broke off (i.e., did not complete key survey and demographic questions). There were no differences in the rates of breakoffs among student, staff, and faculty respondents (Fischer’s Exact Test(n=3922)=2.790, $p=.240$). In response to invitations that had the authoritative subject line, “MSU Vice President for Finance and Operations asks you to take a survey,” all three groups responded at significantly higher rates. That is, the response rate after two contacts for faculty was 7% higher ($p<.01$), for staff it was 4% higher ($p<.10$), and for students it was 2% higher ($p<.01$) for those receiving the authoritative subject line as opposed to those receiving the salient subject line.

Marginal and Interaction Effects

In order to better understand the relative impact of the different invitation elements on response rates, we developed logit regression models of response. These models were able to draw on the information that we had about all subjects receiving an invitation to participate in our study. The limited information that the registrar provided to us about all the members of our sample included the prospective subjects' "local address" and whether a subject is faculty, student or staff. We also knew the design characteristics of the invitations sent to every member of our 15,648 sample list. Thus, the dependent variable for our logit is a binary variable indicating response (coded as 1) or no response (coded as 0), and the independent variables are the design elements and the address variables. Separate models are estimated for faculty, students and staff (a pooled model was rejected using a likelihood ratio test with $p<.01$).

The results of the logit analyses for the first two waves of response data are presented in Table 4. The reported output includes odds ratios, marginal effects, and p-values for the marginal effects. We report the odds ratios because they are so commonly reported as output from logistic regression models. Odds ratios greater (less) than one indicate that a variable increases (decreases) the odds of a response versus non-response. Although common in the literature, in our context the odds ratios are difficult to directly interpret since they are ratios of ratios. Alternatively, the marginal effects are easily interpreted as the change in predicted response rate due to a one unit change in a variable. Since our treatment variables are binary, the reported marginal effects represent the additional predicted response rate with a design element compared to the predicted response rate without the design element. Because they are easily understood as changes in response rates, we focus our discussion on the marginal effects of the faculty, staff, and student logit models.

[Insert Table 4 about here]

The logit analysis allows us, among other things, to directly measure the marginal effect of the various invitation design elements and combination of design elements (interaction effects). Because the design elements were dichotomous, one of each 'pair' of design elements is represented by a dummy variable-Postcard, Long text, Top URL, Ten Minute, and VP Subject. Using interaction terms, we report the latter four effects conditional on either the postcard or e-mail treatment. Table 4 also presents results for variables based on the university's 'local' mailing address of record for subjects. Campus address accounts for subjects with campus dorms, offices, etc. as their local mailing address; Local (<10 mi) addresses are those subjects providing a preferred mailing addresses that is off-campus but close by; and Non-local address captures those subject who provided to MSU a 'local' mailing address greater than 10 miles away. The 10 mile radius was selected to capture the main urban and suburban areas around the campus. Subjects with non-local addresses may be university extension agents, research station personal, and satellite campus faculty.

Campus/Local and Mode. The subjects received either a postcard or an e-mail invitation for the first two waves of contacts. Table 4 shows that subjects who provided the university with a campus address as their mailing address were significantly ($p<.01$) higher response rates than subjects who provided an off campus address greater than 10 miles away–faculty (13%), staff (19%), and students (4%). Likewise, subjects that provided a local address less than 10 miles away from campus had higher response rates for staff (16%, $p<.05$) and students (4%, $p<.01$), but not for faculty.

The address variables were interacted with the postcard invitation mode to test for effects specific to the invitation mode. As Table 4 shows, the significant marginal effect of inviting faculty to participate in a Web survey with a postcard to their campus address was negative (-19%, $p<.01$). Taking into account the positive marginal effect of inviting "campus" faculty to participate in the survey (13%), the net marginal response rate effect of inviting on-campus faculty by postcard is about -5%. In other words, using e-mail with this group would result in increased response rates. Whether the invitation was send by e-mail or postcard did not affect response from campus or local students and staff. The small number of postcard invitations sent to staff with off-campus addresses were associated with increased response (31%, $p<.01$). Students sent postcard invitations to non-local addresses resulted in decrease response rates (-6%, p<01). These results underscore the finding that for some populations, such as local staff and students in our case, Web survey invitations by e-mail or postcard result in substantially similar response rates. Conversely, for other populations, such as faculty using campus mail, there is an increased response to Web survey invitations delivered by e-mail.

Invitation Text Length. As Table 3 indicated, the long text version of the invitations seem to make a positive difference on some response rates. Table 4 reports the logit model results further exploring the marginal impact of text length and invitation mode effects. Postcard Long text results show that the length of text used in the postcard invitation had no significant effect on the response rates for faculty, staff, and students. However, E-mail Long text results show significant and positive response rate effects for the long text e-mail invitations sent to faculty (7%, $p<.05$) and staff (4%, $p<.10$). The increased length of text in the e-mail invites to faculty and staff increased their response. For students however, the text length of the e-mail invitations did not make any difference.

URL Location. The results of the logit model show the unsurprising result that URL location (top versus bottom) does not matter for response to postcard invitations. The subjects receive their postcard, read it, and decide whether to logon to the Web survey. To do so, they need to key in their logon information (URL) from the postcard, and it appears not to matter if this information is at the top or bottom of the postcard. The results do show the URL location does matter for e-mail Web survey invitations. As Table 4 illustrates, 6% fewer faculty responded to the e-mail invite with the URL at the top ($p<.05$) and 2% fewer students responded to e-mail invitations with the URL at the top ($p<.01$). Since there was no significant difference in the response of staff to e-mail with URL at top or bottom, it seems clear that placing the URL link at the bottom of an e-mail invitation will increase response to Web survey invitations.

Survey Effort Estimate. The marginal effect on response rates associated with informing subjects that the survey would take about 10 minutes (in contrast to less than 30 minutes) was examined by mode. Interestingly, the different effort estimates did not result in significantly different response from faculty, staff, and students expect in one instance. The response rate of students who received e-mail invitations to participate in a Web survey that was estimated to take about 10 minutes was 3% larger ($p<.01$) than for an e-mail indicating less than 30 minutes. Unlike faculty and staff who might either decide to participate or not participate in a survey for other reasons and who may discount or dismiss survey effort estimates, the students appear to take seriously the survey effort (time) estimates in e-mail invitations.

Subject Line. As the results from Table 3 indicated, the authoritative subject line in Web survey invitations appears to significantly increase Web survey response. The logit model results in Table 4 shed some additional light of the study results. First of all, use of the authoritative subject line (VP Subject) is always positive and significantly so for all groups receiving the postcard invitations and for faculty receiving e-mail invitations. That is, postcard invitations that used a subject line that said, "MSU Vice President for Finance and Operations asks you to take a survey" increased response from faculty (10%, $p<.10$), staff (7%, $p<.10$) and students (4%, $p<.01$). The authoritative e-mail subject line did not seem to matter for staff and students after two waves of contacts but was associated with about 7 percent additional response rate from faculty ($P<.10$).

Response after Three Contacts

Table 5 presents the aggregate response rate results by treatment and group after all three waves of contacts were completed. Recall that roughly one-third of subjects received a postcard invitation for their first two contacts and an e-mail invitation as their third and final reminder. The first thing to note is that the aggregate response rates increased across the board as a result of the third contact. At the same time, several relative differences also become apparent. First, the mode difference observed for staff after two waves of contacts disappears after they receive a third contact, yet a mode effect now emerges for students with the postcard/e-mail combination yielding a significantly higher response rate than the e-mail only group. These results indicate the relative attractiveness of multiple contacts and the use of mixed invitation modes. Also, the significance of response rate differences associated with URL location for students diminishes after the additional e-mail reminder. Staff also evidenced a significant increase in response rates to the longer invitations and the authoritative subject lines after the third contact.

DISCUSSION

The results suggest that, even without using an e-mail invitation, one may be able to obtain Web survey response rates ranging from 14% to 32% by using postcards to invite subjects to participate in a Web survey. This is an especially promising result for situations where researchers may want to take advantage of a Web platform for implementing a survey, but lack an electronic means of contacting a population. Postcards are less expensive than letters, but they do not permit the enclosure of incentives (e.g., $1.00) typically recommended as an invitation design element for increasing mail survey response. However, the postcard invitations in this study (like the e-mail invitations) did offer respondents an opportunity to be included in a raffle to win a $200 [university] computer store gift certificate. We did not test a postcard versus letter invitation to participate in a Web survey in our study.

The prevailing wisdom that e-mail invitations are always preferred over other forms of invitation to participate in a Web survey (e.g., letters or postcards) is not fully supported here. The results show that invitation mode, all else equal, results in different response rates for students, staff, and faculty. Email invitations appear to yield higher response than postcard invitations for faculty while postcard invitations resulted in comparable response results to e-mail invitations for the others after three waves of contacts. The results especially for student response reveal that mixed-method recruitment (use of both postcard and e-mail invitations) worked better than e-mail invitations alone. It may be that, given the level of SPAM received by college students these days, the novelty and legitimacy of a postcard invitation may serve to counter the advantages of easy access to the Web survey offered by the e-mail invitation. This is an intriguing finding deserving of more research.

The response effects of the other invitation design elements differed across the populations. Students did not always respond to the embedded invitation design elements in the same ways as faculty or staff. It does seem clear from the results that three waves of invitations, all else equal, resulted in greater response from all of the population segments as compared with response rates after two invitations. Subjects initially sent postcard invitations did seem to have a large third wave increase in their response rates after the e-mail (third) invitation as compared to the increase in response from the third wave of e-mail invitations to those who had previously received e-mails.

The URL placement did not seem to matter to subjects except for those students and faculty receiving an e-mail invitation. One post hoc explanation may be that placing the URL lower in the invitation may encourage recipients to read the entire message, which may increase the legitimacy of the request. Placing the URL at the top of the invitation is clearly not better for any of our three groups and placing the URL at the bottom of the invitation was clearly better for faculty and students.

While the conventional wisdom is that shorter is better, this did not prove to be the case for Web survey invitations' estimates of survey effort for faculty and staff. Only those students receiving e-mail invitations with the estimate of "about 10 minutes" responded at a significantly greater rate after three contacts (about 2%, $p<.01$). The fact that this information was in the second paragraph of the invitation may have made it less visible. For this topic and for these populations at least, estimated length of the survey did not have a big effect on response. Shorter estimates may be better for some populations but not all populations and not by much in the aggregate.

The longer invitation text treatments appear to be better than the shorter invitations at increasing response rates for faculty and staff while length of invitation text appears to make no difference in student response rates. This suggests that subjects may prefer more information rather than less about the study, although our treatments tried to control for the same content in both long and short treatments. Therefore, there may be a 'seriousness' or 'importance' signal that a longer text conveys to subjects that accounts for their increased response to the longer invitations. This finding runs counter to the conventional wisdom that shorter invitations are better, and suggests another area for further research.

The results do support the notion that use of an authoritative subject line increases response rates to Web survey invitations. After two contacts, this proved to be the case for faculty, staff, and students who received postcard invitations as well as to faculty receiving e-mail invitations. After three contacts, all three sample populations evidenced significantly greater response rates to invitations that used the authoritative subject line. The authoritative request for input does not depend upon subject matter salience or interest as a motivating factor. Announcing the topic of the survey in the subject line may increase the potential for nonresponse bias by attracting those with greater interest in the topic. Our results suggest that using an authoritative subject line without mention of the topic appears to be effective.

CONCLUSION

Our findings clearly establish that the design of Web survey invitations matters for response rates. Interestingly, despite strong priors that telling subjects that the survey would take less time would increase response, there were no significant differences in the response rate for faculty and staff for the different survey effort treatments. Furthermore, longer invitation text resulted in increased response rates for faculty and staff for the e-mail mode and no significant differences for students. Placing the URL at the bottom of Web survey invitations rather than at the top seems to be more effective in encouraging response. The study results support the use of an authoritative subject line in Web survey invitations. However, it remains uncertain what elements of such an authoritative subject line are most persuasive and whether an increased use of such subject lines might diminish their positive impact on response rates. Compared to e-mail invites, postcards were detrimental to response rates for the faculty group although a mixed postcard/e-mail approach after three waves appeared beneficial for student response. Nevertheless, the study results also demonstrate that using postcards to invite participation in a Web survey works fairly well despite the fact that recipients of the postcards had to go to a computer to manually logon to the survey using the unique URL assigned to them. The findings reveal that the effect of many of the invitation design elements depends on the invitation mode. Finally, the considerable variation in some of the findings between faculty, staff and students at the same institution suggest caution in generalizing findings from studies done among one type of population to others. However, our study suggests that more research effort should be paid to the content, form, and method of delivery of the invitation to participate in a Web survey.

Literature Cited

American Association for Public Opinion Research (AAPOR). 2006. Standard definitions: Final dispositions of case codes and outcome rates for surveys. http://www.aapor.org/ethics/stddef.html#mail.

Bosnjak, M., Neubarth, W., Couper, M.P., Bandilla, W., and Kaczmirek, L. 2008. Prenotification in Web-Based Access Panel Surveys: The Influence of Mobile Text Messaging versus E_Mail on Response Rates and Sample Composition. *Social Science Computer Review* 26(2): 213-223.

Cobanoglu, C., B. Warde, and P. J. Moreo. 2001. A comparison of mail, fax, and web-based survey methods. *International Journal of Market Research* 43: 441-452.

Cook, C., F. Heath, and R. L. Thompson. 2000. A meta-analysis of response rates in web- or Internet-based surveys. *Educational and Psychological Measurement* 60: 821-826.

Couper, M. P. 2000. Web surveys: a review of issues and approaches. *Public Opinion Quarterly* 64: 464-494.p

Couper, M. P., M. W. Traugott, and M. J. Lamias. 2001. Web survey design and administration. *Public Opinion Quarterly* 65: 230-253.

Couper, M.P., Kapteyn, A., Schonlau, M. & Winter, J. 2007. Noncoverage and Nonresponse in an Internet Survey. *Social Science Research* 36(1), 131-148.

Couper, M.P., Tourangeau, R., Conrad, F.G. & Crawford, S.D. 2004. What They See Is What We Get - Response Options For Web Surveys. *Social Science Computer Review* 22(1), 111-127.

Crawford, S.D., Couper, M.P., and Lamias, M. 2001.Web Surveys: Perceptions of Burden. *Social Science Computer Review* 19 (2): 146-162.

Crawford, S.D., McCabe, S.E., Saltz, B., Boyd, C.J., Freisthler, B., and Paschall, M.J. 2004.Gaining Respondent Cooperation in College Web-Based Alcohol Surveys: Findings from Experiments at Two Universities. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ, May.

De Leeuw, E.D., Callegaro, M., Hox, J.J., Korendijk, E., and Lensvelt-Mulders, G. 2007. The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis. *Public Opinion Quarterly*, 71 (3): 1-31.

Denscombe, M. 2006 Web-Based Questionnaires and The Mode Effect - An Evaluation Based on Completion Rates and Data Contents of Near-Identical Questionnaires Delivered in Different Modes. *Social Science Computer Review* 24(2), 246-254.

Dillman, D. A. 2007. *Mail and Internet surveys: the Tailored Design Method*. 2nd Ed. New York: Wiley.

Dillman, D. A., J. R. Clark, and M. A. Sinclair. 1995. How prenotice letters, stamped return envelopes, and reminder postcards affect mailback response rates for census questionnaires. *Survey Methodology* 21:1-7.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R. & Kwan, I. 2002. Increasing Response Rates to Postal Questionnaires: Systematic Review. *British Medical Journal* 324(7347): 1183-1185.

Evans, J.R. & Mathur, A. 2005. The Value of Online Surveys. *Internet Research-Electronic Networking Applications and Policy* 15(2), 195-219.

Galesic, M. 2006. Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics*, 22 (2): 313-328.

Groves, R. M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Gueguen, N. and Jacob, C. 2002. Solicitation By E-mail and Solicitor's Status: A Field Study of Social Influence on the Web. *Cyberpsychology & Behavior* 5(4), 377-383.

Harmon, M.A., Westin, E.C., and Levin, K.Y. (2005), "Does Type of Pre-Notification Affect Web Survey Response Rates?" Paper presented at the annual conference of the American Association for Public Opinion Research, Miami Beach, May.

Heerwegh, D. and Loosveldt, G. 2002. Web Surveys: The Effect of Controlling Survey Access Using PIN Numbers. *Social Science Computer Review*, 20 (1): 10-21.

Heerwegh, D. and Loosveldt, G. 2003. An Evaluation of the Semiautomatic Login Procedure to Control Web Survey Access. *Social Science Computer Review*, 21 (2): 223-234.

Hembroff, L.A., Rusz, D., Rafferty, A., McGee, H., and Ehrlich, N. (2005), "The Cost-Effectiveness of Alternative Advance Mailings in a Telephone Survey." *Public Opinion Quarterly*, 69 (2): 232-245.

Huang, H.M. 2006. Do Print and Web Surveys Provide the Same Results? *Computers in Human Behavior* 22(3), 334-350.

Joinson, A.N. and Reips, U.D. 2007. Personalized Salutation, Power of Sender and Response Rates to Web-Based Surveys. *Computers in Human Behavior* 23(3), 1372-1383.

Joinson, A.N., Woodley, A. and Reips, U.D. 2007. Personalization, Authentication and Self-Disclosure in Self-Administered Internet Surveys. *Computers in Human Behavior* 23(1), 275-285.

Kaplowitz, M.D., Hadlock, T.D. & Levine, R. 2004a. A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly* 68(1), 94-101.

Kaplowitz, M.D., Lupi, F., and Hoehn, J. 2004b. Multiple-Methods for Developing and Evaluating a Stated Choice Survey to Value Wetlands. In *Methods for Testing and Evaluating Survey Questionnaire,* Eds. S. Pressor, J. M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer. Hoboken, NJ: John Wiley and Sons, pages 503-524.

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S. & Schutz, A. 2007. Compensating For Low Topic Interest and Long Surveys - A Field Experiment on Nonresponse in Web Surveys. *Social Science Computer Review* 25(3), 372-383.

Porter, S.R. and Whitcomb, M.E. 2005. E-mail Subject Lines and Their Effect on Web Survey Viewing and Response. *Social Science Computer Review* 23(3), 380-387.

Porter, S.R., and Whitcomb, M.E. 2007.Mixed-Mode Contacts in Web Surveys: Paper is not Necessarily Better. *Public Opinion Quarterly* 71 (4): 635-648.

Ray, N.M. and Tabor, S.W. 2003 Cybersurveys Come of Age. *Marketing Research* 15, 32 - 37.

Schaefer, D. R., and D. A. Dillman. 1998. Development of a standard E-mail methodology. *Public Opinion Quarterly* 62: 378-397.

Schaefer, D.R. & Dillman, D.A. (1998) Development of a Standard E-mail Methodology - Results of an Experiment. Public Opinion Quarterly 62(3), 378-397.

Sills, S. J., and C. Song. 2002. Innovations in survey research: an application of web-based surveys. *Social Science Computer Review* 20: 22-30.

Trouteaud, A.R. 2004. How You Ask Counts: A Test of Internet-Related Components of Response Rates to a Web-Based Survey. *Social Science Computer Review* 22(3): 385-392.

Witte, J.C., Amoroso, L.M. & Howard, P.E.N. 2000. Research Methodology - Method and Representation in Internet-Based Survey Tools - Mobility, Community, and Cultural Identity in Survey2000. *Social Science Computer Review* 18(2), 179-195.

Table 1: Invitation Experimental Treatments

Treatment	Alternatives	Description
Mode	Postcard	Up to 2 postcard invitation and final e-mail
	E-mail	Up to 3 e-mail invitations
Length	Long	Long text wording (182 words)
	Short	Short text wording (80 words)
URL	Top	URL link at top of invitation
	Bottom	URL link at bottom of invitation
Effort	Low	"The survey will take about 10 minutes"
	High	"The survey will take less than 30 minutes"
Subject Line	Authority	"MSU Vice President for Finance and Operations asks you to take a survey"
	Subject Salience	"Take an MSU survey on campus environmental stewardship"

Table 2: AAPOR Minimum Response Rates (RR1) for Survey Populations

Group	Addressees	Completes	RR1
Total	15,648	3,896	24.9%
Faculty	1,488	563	37.8%***
Staff	2,064	883	42.8%***
Students	12,096	2,450	20.3%***

*** Sub-group response rates significantly differ, $\chi^2(2,n=15648)=630.85$, $p=0.000$

Table 3: Response Rate (RR1) Differences by Treatment and Group after Two Contacts †

		Faculty		Staff		Students	
Treatment	**Variable**	***N***	**RR1**	***N***	**RR1**	***N***	**RR1**
Mode	Postcard	494	20.9%***	687	32.0%*	4036	14.0%
	E-mail	987	**32.6%*** **	1376	**36.3%***	8072	15.0%
Length	Long Text	740	**31.4%****	1032	**36.7%***	6056	15.0%
	Short Text	741	26.0%**	1031	33.0%*	6052	14.4%
URL	Top URL	741	26.3%**	1031	35.0%	6056	13.9%**
	Bottom URL	740	**31.1%****	1032	34.7%	6052	**15.5%****
Effort	Ten Minute 10min	740	28.9%	1032	33.5%	6054	**15.8%*****
	Thirty Minute	741	28.5%	1031	36.2%	6054	13.6%***
Subject	VP Subject	740	**32.4%*****	1032	**36.7%***	6054	**15.8%*****
	Enviro. Subject	741	25.0%***	1031	33.0%*	6054	13.6%***

† Chi-Square Test of Significant Differences: *** = p < .01, ** = p < .05, * = p<.10

Table 4: Odds Ratios and Marginal Effects for Logit Models Relating Response to Invitation Design Elements. †

	Faculty			Staff			Students		
Variable	**Odds Ratio**	**Marginal Effect**	**p-value**	**Odds Ratio**	**Marginal Effect**	**p-value**	**Odds Ratio**	**Marginal Effect**	**p-value**
Campus address	**2.21**	**0.134**	**0.002**	**2.73**	**0.194**	**0.000**	**1.34**	**0.035**	**0.003**
Local (<10 mi) address	2.00	0.154	0.224	**1.99**	**0.164**	**0.046**	**1.44**	**0.044**	**0.000**
Postcard to Campus address	**0.35**	**-0.187**	**0.000**	0.80	-0.049	0.327	1.07	0.008	0.665
Postcard to Local address	0.64	-0.080	0.471	1.30	0.060	0.537	0.93	-0.009	0.577
Postcard Non-local address	1.33	0.059	0.616	**3.61**	**0.310**	**0.002**	**0.51**	**-0.064**	**0.000**
Postcard Long text	1.03	0.006	0.888	1.13	0.028	0.467	1.05	0.006	0.578
E-Mail Long text	**1.44**	**0.074**	**0.011**	**1.21**	**0.044**	**0.092**	1.05	0.006	0.433
Postcard Top URL	1.06	0.012	0.795	0.94	-0.014	0.701	1.09	0.010	0.351
E-Mail Top URL	**0.73**	**-0.060**	**0.021**	1.01	0.002	0.942	**0.81**	**-0.024**	**0.001**
Postcard Ten Minute	1.40	0.069	0.163	0.85	-0.035	0.332	0.97	-0.003	0.779
E-Mail Ten Minute	0.90	-0.020	0.451	0.97	-0.006	0.820	**1.30**	**0.032**	**0.000**
Postcard VP Subject	**1.61**	**0.101**	**0.050**	**1.34**	**0.067**	**0.086**	**1.34**	**0.037**	**0.003**
E-Mail VP Subject	**1.43**	**0.072**	**0.013**	1.20	0.041	0.113	1.10	0.011	0.143

† Since the treatment variables are dichotomous, marginal effects are computed using the differences in predicted probabilities of response with the respective variable set at one and zero where all other variables are evaluated at their mean. The reported p-values are for the marginal effects computed using the delta method.

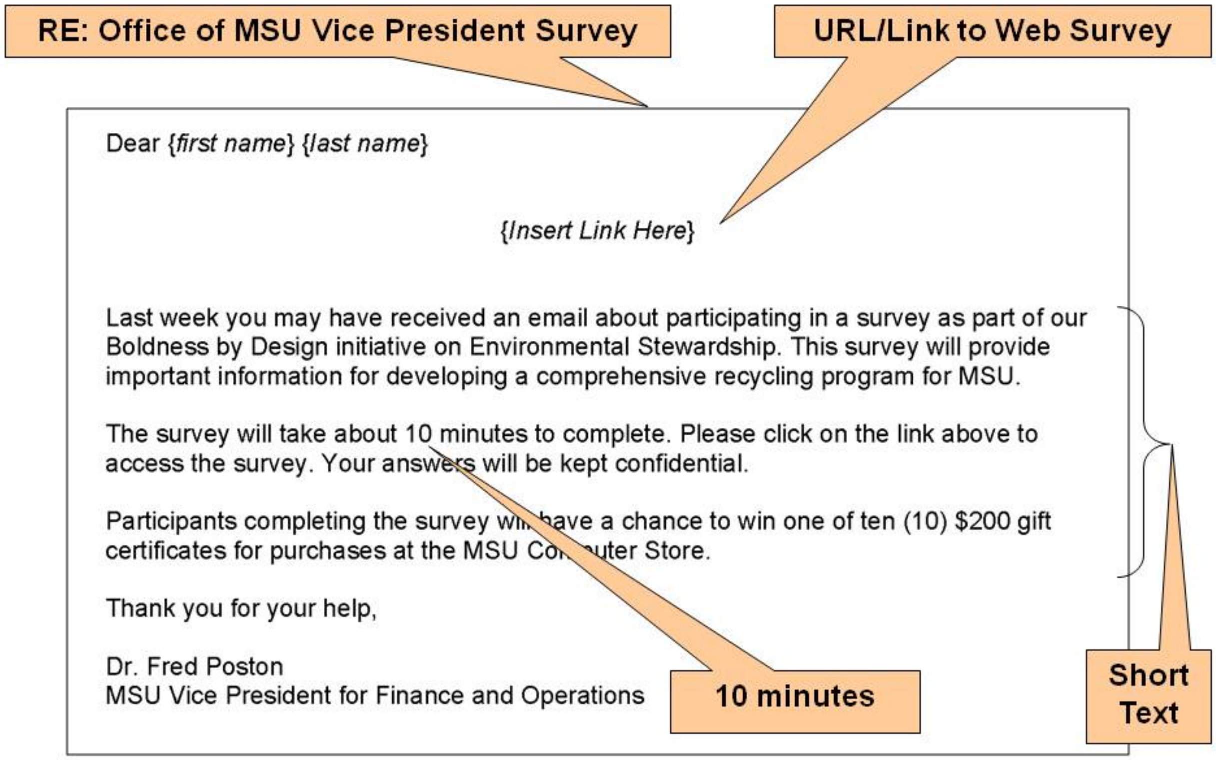
Table 5: Response Rate (RR1) Differences by Treatment and Group after Three Contacts †

		Faculty		Staff		Students	
		Sample size	**RR1**	**Sample size**	**RR1**	**Sample size**	**RR1**
Mode	postcard	494	33.4%***	687	43.2% [1]	4036	**21.8%*** [2]
	e-mail	987	**40.3%*****	1376	42.6% [1]	8072	19.4%*** [2]
Length	short	741	35.2%**	1031	40.6%** [2]	6052	20.5%
	long	740	**40.8%****	1032	**45.0%**** [2]	6056	29.9%
URL	top	741	35.1%**	1031	42.0%	6056	19.5%* [1]
	bottom	740	**40.9%****	1032	43.6%	6052	**20.9%*** [1]
Effort	~ 10min	740	38.9%	1032	42.3%	6054	**21.3%*****
	< 30min	741	37.1%	1031	43.3%	6054	19.2%***
Subject	Vice President Survey	740	**42.8%*****	1032	**45.6%***** [2]	6054	**21.8%*****
	Environmental Survey	741	33.2%***	1031	40.0%*** [2]	6054	18.7%***

† Chi-Square Test of Significant Differences: *** = p < .01, ** = p < .05, * = p < .10

[1] Decreased in significance as compared to results after two contacts

[2] Increased in significance as compared to results after two contacts



Dear {*first name*} {*last name*}

{*Insert Link Here*}

Last week you may have received an email about participating in a survey as part of our Boldness by Design initiative on Environmental Stewardship. This survey will provide important information for developing a comprehensive recycling program for MSU.

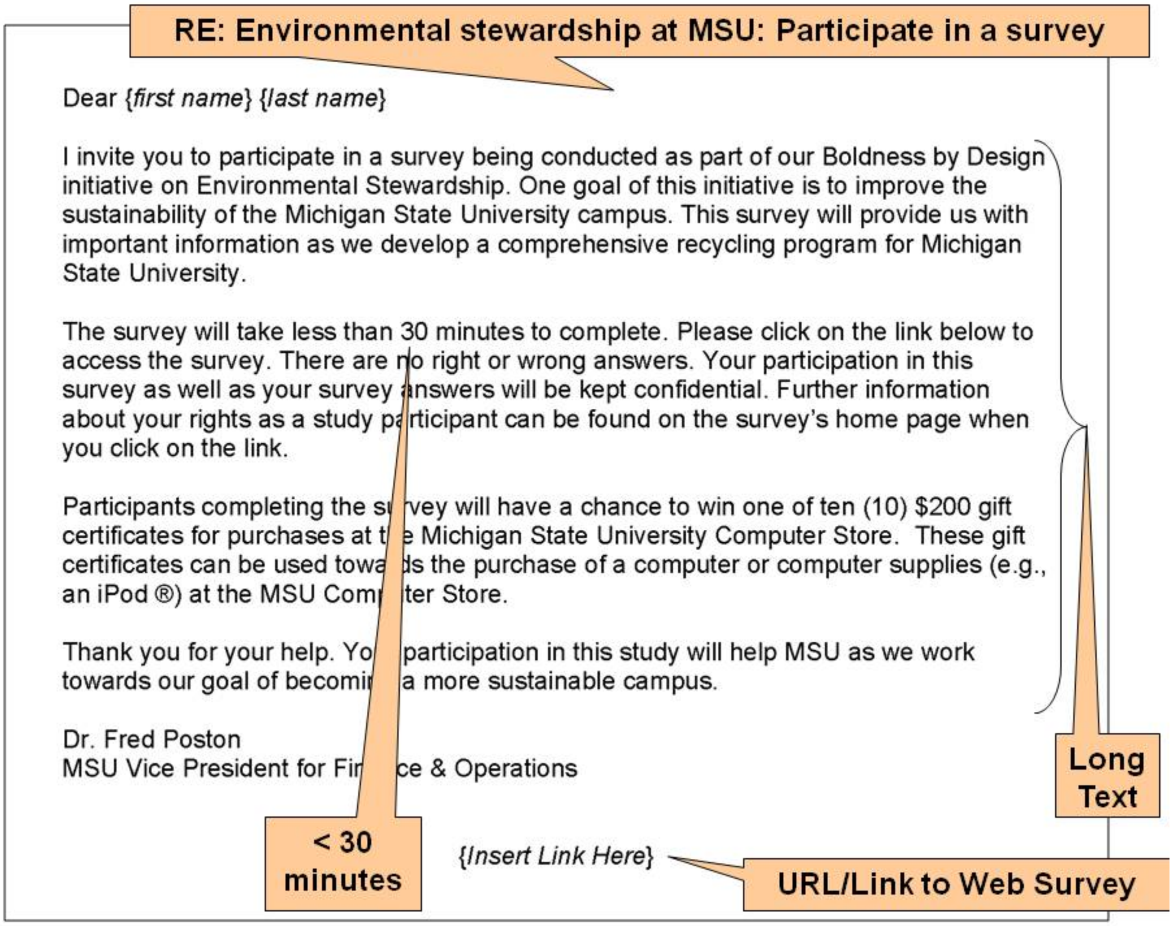
The survey will take about 10 minutes to complete. Please click on the link above to access the survey. Your answers will be kept confidential.

Participants completing the survey will have a chance to win one of ten (10) $200 gift certificates for purchases at the MSU Computer Store.

Thank you for your help,

Dr. Fred Poston
MSU Vice President for Finance and Operations

Figure 1: Short text invite with 10 minute effort, top URL, and VP subject line design


RE: Environmental stewardship at MSU: Participate in a survey

Dear {*first name*} {*last name*}

I invite you to participate in a survey being conducted as part of our Boldness by Design initiative on Environmental Stewardship. One goal of this initiative is to improve the sustainability of the Michigan State University campus. This survey will provide us with important information as we develop a comprehensive recycling program for Michigan State University.

The survey will take less than 30 minutes to complete. Please click on the link below to access the survey. There are no right or wrong answers. Your participation in this survey as well as your survey answers will be kept confidential. Further information about your rights as a study participant can be found on the survey's home page when you click on the link.

Participants completing the survey will have a chance to win one of ten (10) $200 gift certificates for purchases at the Michigan State University Computer Store. These gift certificates can be used towards the purchase of a computer or computer supplies (e.g., an iPod ®) at the MSU Computer Store.

Thank you for your help. Your participation in this study will help MSU as we work towards our goal of becoming a more sustainable campus.

Dr. Fred Poston
MSU Vice President for Finance & Operations

{*Insert Link Here*}

Figure 2: Long text invite with < 30 minute effort, bottom URL, and environmental subject design

Incorporating Random Coefficients and Alternative Specific Constants into Discrete Choice Models: Implications for In-Sample Fit and Welfare Estimates *

H. Allen Klaiber
Department of Agricultural and Resource Economics
North Carolina State University

Roger H. von Haefen
Department of Agricultural and Resource Economics
North Carolina State University

This Draft: March 2008

Abstract

In recent years, several innovative econometric methods have been employed in non-market valuation applications of discrete choice models. Two particularly attractive methods are random parameters (which introduce more plausible substitution patterns) and alternative specific constants (which control for unobserved attributes). In this paper, we investigate the properties of these methods along several dimensions. Across three recreation data sets, we consistently find large improvements in model fit arising from the inclusion of both methods; however, these gains often come concomitant with significant degradations in-sample trip predictions. We then show how poor in-sample predictions correlate with welfare estimates. Using econometric theory and Monte Carlo evidence, we illuminate why these perverse findings arise. Finally, we propose and empirically evaluate four 'second-best' modeling strategies that attempt to correct for the poor in-sample predictions we find in our applications.

Keywords: Discrete Choice, Recreation Demand, Revealed Preference, Stated Preference, Welfare Estimation, Alternative Specific Constants, Random Parameters

---

*Early draft that is currently under major revision. Comments welcome, but please do not cite. We thank George Parsons, Bill Breffle, Dan Phaneuf, Chris Timmins, and W2133 Annual Meeting participants for helpful comments.

Section I. Introduction

Discrete choice models have become one of the most frequently used modeling frameworks for recreation demand and locational equilibrium models (Murdock, 2006; Bayer and Timmins, 2007). Within the framework, two econometric innovations that applied researchers are using with increasing regularity are random coefficients (McFadden and Train, 2000) and the inclusion of alternative specific constants (Berry, 1994). Random coefficients are an attractive mechanism for relaxing the restrictive implications of the independence of irrelevant alternatives (IIA), thus introducing more plausible substitution patterns. Including a full set of alternative specific constants allows the analyst to control for unobserved attributes that may be correlated with observed attributes.

In applications of these modeling innovations to discrete choice models, researchers have found that they generate substantial and statistically significant improvements in fit (von Haefen and Phaneuf, 2008; Murdock, 2006). In an empirical investigation of three recreation data sets, we also find large gains in model fit. However, we also find that the models with alternative specific constants and random coefficients often fail to replicate the in-sample aggregate visitation patterns implied by the data. This empirical regularity generates important implications for the credibility of welfare analysis – why should one believe welfare measures derived from models that cannot replicate in-sample aggregate choice behavior?

Our goal in this paper is to shed light on the counterintuitive empirical regularity of improved statistical fit combined with poor in-sample prediction. We begin by documenting this phenomenon with three recreation data sets that have been used in previously published research. Two of the three applications combine revealed and stated preference (RP-SP) to identify all demand parameters (Adamowicz et al., 1997; Haener et al., 2001) as previously done by von Haefen and Phaneuf (2008). The other exploits only revealed preference (RP) data (Parsons et al., 1999) and uses a variation of the two-step estimator proposed by Berry, Levinsohn, and Pakes (2004) and used recently by Murdock (2006) in the recreation context. With all three data sets, we find the introduction of random coefficients and alternative specific constants (ASCs hereafter) substantially and significantly improves statistical fit as measured by the log-likelihood. We also find that in-sample trip predictions often (but not uniformly) deteriorate with these richer empirical specifications, and we document how these poor predictions correlate with welfare estimates for a range of policy scenarios.

We then explore why the poor predictions arise in practice. Here we use theoretical results from Gourieroux, Monfort, and Trognon (1984) about the properties of the linear exponential family of distributions as well as some Monte Carlo findings. The upshot of our discussion is that: 1) fixed coefficient logit models with a full set of ASCs will generate in-sample trip predictions for each alternative that *perfectly* match the data, and 2) random coefficient logit models with or without ASCs may not predict perfectly in-sample, but should generate reasonably close predictions if the analyst has correctly specified the underlying data generating process. An implication of this finding is that the poor in-sample predictions that we find in our three applications arise because of model misspecification. Thus, logit models with random coefficients and ASCs fit the data better than models without these econometric innovations, but they nevertheless fail to account for important features of the data.

We conclude by exploring a number of 'second best' strategies for dealing with poor in-sample predictions. These range from: 1) abandoning random coefficient specifications and using fixed coefficient models with ASCs that generate perfect in-sample predictions; 4) using less-efficient non-panel random coefficient models that, as we demonstrate, generate more plausible in-sample predictions; 3) using the Berry (1994) contraction mapping or maximum penalized likelihood (Montricher et al., 1975, Silverman, 1982; Huh and Sickles, 1994; Shonkwiler and Englin, 2005) with ASCs to force the in-sample predictions to match the data perfectly; and 4) conditioning on observed choice in the construction of welfare measures following von Haefen (2003). Our preliminary results suggest that each of these strategies is effective in terms of generating plausible in-sample predictions but they differ considerably in terms of their implications for statistical fit.

The paper proceeds as follows. The next section documents the performance of fixed and random coefficient logit models with and without a full set of ASCs with three recreation data sets. Section III explores the factors that give rise to the perverse empirical findings reported in the previous section using econometric theory and a set of Monte Carlo simulations. Section IV investigates a number of 'second best' empirical strategies that applied researchers may find attractive in future applications. We then conclude with some final observations and recommendations.

Section II. Nature of the Problem

We begin by illustrating the poor in-sample prediction problem that serves as the motivation for this research. To demonstrate that this problem is not an idiosyncratic feature associated with a single data set, we consider three recreation data sets that researchers have used in previously published studies. The first data set was first used by Adamowicz et al. (1997) and consists of both revealed preference (RP) and stated preference (SP) choice data for moose hunting in the Canadian province of Alberta. The RP data consists of seasonal moose hunting trips for 271 individuals to 14 wildlife management units (WMUs) throughout Alberta in 1993. The SP data consists of 16 choice experiments that were generated with a blocked orthogonal, main effects design. All eleven site attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. The second data set was first used by Haener et al. (2001) and also consists of combined RP-SP data for Canadian moose hunting. This data source, however, was collected in the neighboring province of Saskatchewan in 1994. The RP data consists of seasonal moose hunting trips for 532 individuals to 11 wildlife management zones (WMZs) throughout Saskatchewan. The SP data consists of 16 choice experiments that were generated with a blocked orthogonal, main effects design. All nine attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. As discussed in von Haefen and Phaneuf (2007), the fusion of RP and SP data is attractive in both data environments because the inclusion of a full set of ASCs confounds identification of the site attribute parameters given the relatively small number of sites in each application.

For both data sets, we control for differences in scale across RP and SP data sources and use empirical specifications, estimation strategies, and welfare scenarios that match those used by von Haefen and Phaneuf (2008).

The third data set we consider looks at Mid-Atlantic beach visitation and was first used by Parsons et al. (1999). This data set consists of seasonal trip data to 62 ocean beaches in 1997 for 375 individuals. For each beach, we observe 14 site characteristic variables plus we construct individual-specific travel costs based on each recreator's home zip code. Because we use only RP data with this application, we use a two-step estimation strategy for those models that includes a full set of ASCs (Berry, Levinsohn, and Pakes, 2004; Murdock, 2006). For the results reported in Table 1, our two-step estimator differs from previous two-step estimators in the following way. Similar to Murdock, we use maximum likelihood techniques in the first step to estimate the travel cost parameter and a full set of ASCs that subsume all 14 site characteristics that do not vary over individuals (note: we do not include any demographic interactions in this model because preliminary testing suggested that they did not improve model fit). In contrast to Murdock, our first step estimator does not employ the Berry (1994) contraction mapping algorithm, an issue we return to in a later section. Thus, our first step estimator relies entirely on traditional maximum likelihood techniques, not the combination of maximum likelihood and Berry contraction mapping techniques that Murdock employs.[23] Our second-stage estimator is identical to Murdock's approach in that we regress the estimated ASCs from the first stage on the 14 site characteristics and a

[23] Using traditional maximum likelihood estimation techniques without the Berry contraction mapping is feasible in our application due to the relative small number of sites in the Mid-Atlantic data set. However, computational tractability requires the use of the Berry contraction mapping in random coefficient applications with many sites.

constant term. Importantly, this approach assumes that the unobserved site attributes are uncorrelated with observed site attributes.

Table I summarizes our findings.[24] All random coefficient models assume that the main effects for the site attributes (excluding travel cost) are normally distributed with no correlations. In on-going work, we are exploring truncated normal and latent class mixing distributions. Arrayed across columns 2-5 are results from four alternative specifications that differ in terms of the inclusion/exclusion of ASCs and random coefficients. In particular, column 2 contains results from models with neither ASCs nor random coefficients, column 3's results contain ASCs but no random coefficients, column 4's results contain random coefficients but no ASCs, and column 5's results contain both. Note that all random parameter specifications assume that all main effects for the various site attributes vary randomly across the population but are common for a given individual, so we refer to these specifications as 'panel' random coefficient specifications following Train (1998). Beginning first with the Alberta results, we note that relative to our baseline model without ASCs and random coefficients, the addition of these modeling innovations generates substantial improvements in fit. The largest gains seem to come from the addition of random coefficients that introduce correlations across an individual's multiple trips, although likelihood ratio tests suggest that ASCs also improve model fit significantly (p value < 0.0001).

To ascertain how well these models predict aggregate trip taking behavior for each site, we construct the following summary statistic for each model:

---

[24] Parameter estimates are available upon request.

(1)

$$\text{Percentage absolute prediction error} = 100 \times \sum_{i=1}^{J} s_i^S \frac{\text{abs}(s_i^S - s_i^M)}{s_i^S} = 100 \times \sum_{i=1}^{J} \text{abs}(s_i^S - s_i^M),$$

where s_i^S and s_i^M are the in-sample share of trips to site *i* and the model's prediction of the share of trips to site *i*, respectively, and *J* is the number of sites. The prediction error statistic can be interpreted as the share weighted in-sample prediction error for each site and thus can be used to rank order the models in terms of in-sample predictions that match the observed data. Intuitively, a model that can replicate aggregate trip predictions well for each site would generate a low prediction error value, whereas a model with poor in-sample aggregate predictions for each site would score a relatively high value. For the Alberta data, we see that the fixed coefficient specification with ASCs has the lowest prediction error statistic (effectively zero), whereas the random coefficient without ASCs has the highest. Interestingly, the substantially better fitting random coefficient with ASCs model has a prediction error statistic that is similar in magnitude to the more parsimonious fixed coefficient without ASCs specification.

Finally, it is interesting to see how these differences in fit and prediction play out in terms of welfare estimates. We consider two scenarios – a reduction in moose population at WMU #348 and an increase in moose population at WMU #344 – and calculate the partial equilibrium (i.e., ignoring changes in congestion) compensating surplus for both scenarios using the approach first suggested by Train (1998). In addition to point estimates and standard errors for the welfare measures, we also report the percentage in-sample prediction error for those sites directly affected by the different policies. Overprediction of the share of trips to these sites is likely to translate into larger welfare estimates, although variability in parameter estimates and the structure of substitution implied by the different models will also play a significant role. For the moose population reduction scenario, we find a range of point estimates from -$9.47 to -$25.00 with significant variation in these estimates' precision. The fixed coefficient with ASCs specification generates in-sample predictions for trips to WMU #348 that match the data well, whereas the other specifications overpredict trips to WMU #348 and generate larger (in absolute value) welfare estimates. For the moose population increase scenario, we find even larger variation in point estimates ($3.61 to $98.34) with significant variation in precision once again. In general, the smaller estimates correspond to specifications that underpredict the share of trips to WMU #344. Based on these results, we conclude that poor in-sample predictions play a significant role in explaining the variation of welfare point estimates in the Alberta data.

Similar results arise with Saskatchewan moose hunting data and the Mid-Atlantic beach data. With both data sets, adding ASCs and especially panel random coefficients improves statistical fit as measured by the log-likelihoods, but this improvement in fit does not necessarily generate lower prediction errors. The percentage absolute prediction errors for the fixed coefficient with ASCs models is once again near zero, but the percentage absolute prediction errors for the panel random coefficient models (with and without ASCs) are uniformly larger than the fixed coefficient models. For the Saskatchewan data, welfare point estimates and their precision vary significantly across the competing models. The variation in point estimates across the competing models seems to be correlated with the degree to which the models over- or underpredict trips to the affected sites. Finally, there appears to be considerably less variation in welfare point estimates for the Mid-Atlantic data, which may be explained by the fact that the alternative models seem to predict in-sample far better for the Mid-Atlantic data than the Alberta or Saskatchewan data.

In summary, the results in Table 1 suggest a somewhat counterintuitive result – including ASCs and especially random coefficients significantly improve overall statistical fit but do not generate in-sample trip predictions that match the observed data well. Welfare measures seem to be correlated with the degree of over- or underprediction implied by the different specifications, but other factors – parameter estimates, the structure of substitution implied by the models – certainly play a significant role. Overall, the results in Table 1 provide mixed evidence in favor of incorporating random coefficients and ASCs into discrete choice models, and cast doubt on the credibility of welfare estimates from models that predict in-sample poorly.

Section III. What explains these counterintuitive results?

In this section we use econometric theory and results from a Monte Carlo analysis to shed light on the counterintuitive results presented in the previous section. To motivate our main insight here, consider the log-likelihood function for a sample of *N* individuals each making separate choices from *J* alternatives:

$$(2) \qquad \ln L(\beta) = \sum_{i=1}^{N} \sum_{j=1}^{J} 1_{ij} X_{ij}\beta - \ln\left(\sum_{k=1}^{J} \exp(X_{ik}\beta)\right),$$

where 1_{ij} is an indicator function equal to 1 for individual *i's* chosen alternative and zero otherwise. The score condition associated with this log-likelihood is:

(3) $$\frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{N} \sum_{j=1}^{J} X_{ij} \left[1_{ij} - \Pr_i(j \mid \beta) \right] = 0 ,$$

where $\Pr_i(j \mid \beta)$ is the logit probability for individual *i* choosing the *j*th alternative. If a full set of ASCs are included, then

(4) $$X_{ij} = \begin{cases} 1 & \text{if } j \text{ chosen} \\ 0 & \text{otherwise} \end{cases}, \forall j ,$$

and the score conditions associated with the ASCs can be written:

(5) $$\sum_{i=1}^{N} \left[1_{ik} - \Pr_i(k \mid \beta) \right] = 0 \quad \text{or} \quad \frac{1}{N} \sum_{i=1}^{N} 1_{ik} = \frac{1}{N} \sum_{i=1}^{N} \Pr_i(k \mid \beta), \forall k.$$

Equation 5 implies that fixed coefficient logit models with a full set of ASCs will generate in-sample predictions that match the data perfectly, a result that is consistent with our empirical findings in Table 1 and well known in the discrete choice literature (see, e.g., Ben-Akiva and Lerman, 1985).

As Gourieroux, Monfort, and Trognon (1984) have shown, the logit distribution falls within the broad class of distributions known as the linear exponential family of distributions. Other notable examples include the Poisson and normal distributions. What defines this family of distributions is that they are all mean-fitting distributions, implying that with the inclusion of ASCs, predictions from these distributions will match the data perfectly. A notable advantage of using linear exponential distributions in empirical work is that if the analyst has correctly specified the conditional expectation function of the distribution (i.e., its first moment), higher order misspecification will not lead to inconsistent parameter estimates (it will, however, bias standard error estimates, but this problem can be addressed if the analyst uses robust standard errors (White, 1981) instead of traditional standard errors). Thus, if the analyst specifies the first moment correctly, consistent parameter estimates will result. This makes the fixed coefficient logit model with ASCs appealing.

What is important to note, however, is that adding random coefficients to the logit distributions results in a mixture distribution that falls outside the linear exponential family. Random coefficient logit models, regardless of whether ASCs are included, will not necessarily generate in-sample predictions that match the data perfectly. This can be seen by looking at the score conditions for the simulated nonpanel random coefficient models logit model. The simulated likelihood function in this case is:

$$(6) \qquad L(\bar{\beta},\sigma) = \sum_{i=1}^{N}\sum_{j=1}^{J} 1_{ij} \ln\left(\frac{1}{R}\sum_{r=1}^{R} \Pr_i(j \mid \beta_i^r)\right) = \sum_{i=1}^{N}\sum_{j=1}^{J} 1_{ij} \ln\left(\frac{1}{R}\sum_{r=1}^{R} \frac{\exp(X_{ij}\beta_i^r)}{\sum_{k=1}^{J}\exp(X_{ik}\beta_i^r)}\right),$$

where $\beta_i^r = \bar{\beta} + \sigma U_i^r$, $U_i^r \sim N(0,1)$, and the score condition is:

$$(7) \qquad \frac{\partial L(\bar{\beta},\sigma)}{\partial \beta_i^r} = \prod_{i=1}^{N}\left[\frac{\prod_{j=1}^{J}\left[X_{ij}\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j \mid \beta_i^r)(1-\Pr_i(j \mid \beta_i^r)\right]^{1_{ij}}}{\prod_{j=1}^{J}\left[\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j \mid \beta_i^r)\right]^{1_{ij}}}\right] = 0 .$$

With the inclusion of ASCs, this condition does not imply perfect in-sample predictions. Thus, some degree of imperfect in-sample prediction can be expected from random coefficient logit models, but the precise degree will vary across applications.

To assess how well in-sample predictions from estimated logit models will match the data, we conducted an extensive Monte Carlo analysis where we know the underlying data generating process for the simulated data. Knowing the true data generating process allowed us to ascertain the in-sample prediction performance of maximum likelihood estimators when model misspecification is absent. If the in-sample predictions generated from these correctly specified models match the observed data well, then we can conclude that poor in-sample predictions arise due to some form of model specification, and not due to an inherent property of the estimator.

For brevity, we only summarize the main conclusions of our Monte Carlo simulation here and leave for an appendix (to be written at a later date – apologies) the simulation details. Across a number of specifications, we consistently found that the in-sample predictions for panel and non-panel random coefficient models with and without alternative specific constants matched the simulated data very closely. Under none of our simulations did we find the degree of poor in-sample prediction that we observed with the Alberta, Saskatchewan, or Mid-Atlantic data – see Table 1. Based on these findings, we conclude that the poor predictions found in our three applications are a result of model misspecification.

The implications of the above discussion for how analysts should proceed are unclear. If the analyst estimates logit models with random coefficients and finds poor in-sample predictions, the obvious 'first best' solution would be to continue to search for empirical specifications that fit the data well and predict well in sample. In practice, however, finding empirical specifications that satisfy these two criteria will be computationally difficult, time-consuming, and in many cases infeasible. This suggests that 'second best' less demanding approaches that address these two concerns may be attractive alternatives to applied researchers. Perhaps the simplest second best approach would be to estimate a fixed coefficient logit model with ASCs where the in-sample aggregate predictions will match the data perfectly. One limitation with this approach is that it in practice employs models with substitution patterns that are consistent with the independence of irrelevant alternatives (IIA). These restrictive substitution patterns can be partially relaxed by using nested logit models, but the considerably more flexible substitution patterns that come with random coefficient models will not be realized.

Another second best approach involves estimating random coefficient models with ASCs using a contraction mapping (Berry, 1994) that iteratively solves for the ASC values by matching the aggregate model predictions with the data. This algorithm was first used in the industrial organization literature to estimate discrete choice models of product differentiation using aggregate market share data (Berry, Levinsohn, and Pakes, 1995), but Berry, Levinsohn, and Pakes (2004) apply the algorithm to a disaggregate data context. Both of these applications employed generalized method of moments estimation techniques, and it was not until Murdock (2006) that the algorithm was used within a maximum likelihood framework with random coefficients. What is interesting to note, however, is that the use of this algorithm within a maximum likelihood estimation procedure *will not* generate maximum likelihood estimates. For this to be the case, the random coefficient maximum likelihood estimates would have to generate in-sample predictions which match the data precisely, but we showed above that in general this will not be the case. Thus, the estimates that one recovers from using the Berry contraction mapping to estimate random coefficient models within the maximum likelihood estimates are akin to maximum penalized likelihood estimates that Shonkwiler and Englin (2005) and von Haefen and Phaneuf (2003) have previously used. The idea behind maximum penalized likelihood estimation is that one maximizes the likelihood subject to a function that penalizes the likelihood for some undesirable behavior. Random coefficient logit models with ASCs that are estimated within the maximum likelihood framework using the Berry contraction mapping are observationally equivalent to estimating random coefficient logit models with ASCs within the maximum penalized likelihood framework with an infinitely weighted penalty function for poor in-sample predictions. A limitation with this approach is that the asymptotic properties of maximum penalized likelihood estimators are not well understood, but it does directly address the poor in-sample prediction problem. Moreover, due to plateaus and non-concavities in the penalized likelihood function, the choice of starting values and search algorithms can strongly influence the derived estimates.

Two other second best approaches for dealing with poor in-sample predictions involve estimating non-panel random coefficient models with ASCs within the maximum likelihood framework or incorporating observed choice into the construction of welfare measures as suggested by von Haefen (2003). As we demonstrate in the next section, the former approaches sacrifice the efficiency gains (which may be substantial) from introducing correlations across an individual's multiple trips for improved (but not perfect) in-sample predictions. Moreover, it makes estimation more computationally intensive. The idea of incorporating observed choice into welfare measurement construction is attractive because it simulates the unobserved determinants of choice in a way that implies perfect prediction for every observation and then uses the model's implied structure of substitution to ascertain how behavior and welfare change with changes in price, quality, and income. The approach can be used with any set of model estimates, but it does require a somewhat more computationally intensive algorithm for calculating welfare estimates (see von Haefen (2003) for details).

In the next section, we compare the sensitivity of welfare estimates to the use of these four second best strategies that address poor in-sample predictions. Our discussion will focus on the Mid-Atlantic application where all welfare measures have been generated. In future revisions to this paper, we will fill in the missing estimates for the Alberta and Saskatchewan data to see how the approaches fair in these alternative data environments.

Section IV. Sensitivity of Welfare Measures to Alternative Second Best Strategies

The bottom third of Table 2 reports welfare estimates from the Mid-Atlantic beach data for two policy scenarios – lost beach width at all Delaware, Maryland, and Virginia (DE/MD/VA) beaches and the closing of all northern Delaware beaches. We report the log-likelihood values as well as the percentage absolute prediction error for all sites in the first two rows to give the reader a sense of the relative statistical fit and in-sample prediction performance of the competing specifications. We also report unconditional (Train, 199?) and conditional (von Haefen, 2003) welfare measures for both scenarios as well as the percentage prediction error at the sites directly affected by the policy for all specifications.

In general, the results reported at the bottom of Table 2 have a number of qualitative implications, although the reader should interpret these implications cautiously until they have been confirmed with the Alberta and Saskatchewan data. First, all of the second best strategies suggested in the previous section for dealing with poor in-sample predictions – using fixed coefficients and alternative specific constants (column 3), using the Berry contraction mapping (columns 8 and 9), and using non-panel random coefficient specifications with alternative specific constants (columns 6 and 8), as well as incorporating observed choice into welfare measures (the conditional welfare measures in all columns) are effective tools for mitigating this problem. Second, the use of non-panel random coefficients results in a significant loss of statistical fit (compare the log-likelihoods in columns 4 and 5, 6 and 7, and 8 and 9). Because the non-panel random coefficient specifications generate smaller prediction error relative to the panel random coefficient models, there is a significant tradeoff between statistical fit and good in-sample prediction when specifying the correlation structure of random coefficients. Third, using the Berry contraction mapping in estimation modestly degrades statistical fit (compare the log-likelihoods in columns 6 and 8 as well as 7 and 9), but it does improve in-sample predictions, especially when panel random coefficients are used.

In terms of welfare estimates, the results in Table 2 imply that there is little difference between the conditional and unconditional welfare across *all* specifications and scenarios. This result is not surprising because the in-sample trip predictions for the affected sites are generally small. For the lost beach width at DE/MD/VA beaches, we see most of the point estimates are clustered in the range of -\$3.34 to -\$11.76, although the estimates that are based on non-panel random coefficient models with ASCs (columns 6 and 8) are positive in sign. As suggested above, the non-panel random coefficient models fit the data far worse than the panel random coefficient models, and thus we doubt the reliability of these estimates which also have rather large standard errors. For the welfare scenario simulating the closing of northern Delaware beaches, we see a general convergence of estimates between -\$11.92 and -\$23.69. We believe this interval represents a plausible range of welfare estimates that should be sufficiently informative for policy purposes.

One could interpret the results from the Mid-Atlantic data as suggesting that the addition of ASCs and random coefficients has minor effects on policy inference. Indeed, the point estimates for the fixed coefficient model without ASCs are qualitatively similar to the mid-range values for the more complex specifications. Based on the incomplete set of results that are reported in Table 2 for the Alberta and Saskatchewan data, we doubt that this empirical finding will carry over to the other applications where prediction error is more extreme. However, one might conclude from the results presented in Table 2 that simple models that predict reasonably well in-sample might generate welfare estimates that are robust to the inclusion of alternative specific constants and random coefficients.

Section V. Conclusion

Our goal in this research has been threefold: 1) to document the somewhat counterintuitive in-sample prediction problems that arise with random coefficient logit models that include ASCs; 2) to explore the sources of these problems using economic theory and Monte Carlo analysis; and 3) to suggest and evaluate alternative, second best, strategies for dealing with the poor in-sample predictions that researchers might find attractive in future empirical work. Across three data sets, we document that the addition of ASCs and especially panel random coefficients generates significant improvements in statistical fit but do not uniformly improve model prediction. We also show how these poor predictions influence derived welfare estimates, with the degree of under- and overprediction at sites that are directly impacted by the policy being correlated with the magnitude of welfare estimates. We then argue that the fixed coefficient logit model falls within the larger family of linear exponential distributions, and thus the inclusion of a full set of ASCs will generate in-sample trip predictions for each site that match the data perfectly. The introduction of random coefficients, however, results in a mixture distribution that falls outside the linear exponential family and thus will not imply perfect in-sample predictions. Results from an extensive Monte Carlo analysis suggest that the poor in-sample predictions observed in our three applications are likely due to some form of misspecification. To account for these model shortcomings, the analyst may find attractive one of the second best strategies that we empirically evaluate for addressing poor in-sample predictions. Our preliminary empirical results with the Mid-Atlantic data suggest that all of these strategies are effective in controlling for poor in-sample predictions, but the use of non-panel random coefficients significantly degrades model fit and generates perverse signs for some of the policy scenarios. Otherwise, our results suggest that the other second best approaches imply qualitatively similar welfare estimates that fall within a narrow range.

Finally, it is worth stepping back and directly addressing the fundamental question that motivated this research: do random coefficients and alternative specific constants improve welfare analysis? With regard to random coefficients, we believe that the richer substitution patterns implied by random coefficients are quite attractive, but the poor in-sample predictions that often result from these models (especially panel random coefficient versions) need to be addressed in some way. If not, welfare estimates lack credibility. With regard to alternative specific constants, we believe that their ability to control for unobserved attributes that may generate endogenity concerns makes them extremely attractive. One limitation with their inclusion, however, is that one needs either an RP data set with many objects of choice (sites in recreation models, or neighborhoods in locational equilibrium models) or additional SP data to identify the part worths of the different site attributes. When these data are available, we believe that ASCs are an attractive modeling innovation.

References

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams, "Perceptions vs. objective measures of environmental quality in combined revealed and stated preference models of environmental valuation," *Journal of Environmental Economics and Management* 32 (1997): 65-84.

Bayer, P. and C. Timmins, "Estimating equilibrium models of sorting across locations," *Economic Journal* 117(2007): 353-374.

Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis*. MIT Press: Cambridge, 1985.

Berry, S. "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25(2): 242-262.

Berry, S., J. Levinsohn, and A. Pakes, "Automobile prices in equilibrium," *Econometrica* 63(1995): 841-890.

Berry, S., J. Levinsohn, and A. Pakes, "Differentiated products demand systems from a combination of micro and macro data: the new car market," *Journal of Political Economy* 112(2004): 68-105.

Gourieroux, C., A. Monfort, and A. Trognon, "Pseudo-maximum likelihood methods: application to Poisson models," *Econometrica* 52(1984): 701-720.

Haener, M., P. Boxall, and W. Adamowicz, "Modeling recreation site choice: Do hypothetical choices reflect actual behavior?" *American Journal of Agricultural Economics* 83(2001): 629-642.

Huh, K., R. Sickles, "Estimation of the duration model by non-parametric maximum likelihood, maximum penalized likelihood and probability simulators," *Review of Economics and Statistics* 76(1994): 683–694.

McFadden, D. and K. Train, "Mixed MNL of discrete response," *Journal if Applied Econometrics*, 15(2000): 447-470.

Montricher, G., R. Tapia, J. Thompson. "Nonparametric maximum likelihood estimation of probability densities by penalty function methods," *Annals of Statistics* 3(1975): 1329–1348.

Murdock, J., "Handling unobserved site characteristics in random utility models of recreation demand," *Journal of Environmental Economics and Management* 51(2006): 1-25.

Parsons, G., D.M. Massey, and T. Tomasi, "Familiar and favorite sites in a random utility model of beach recreation," *Marine Resource Economics*, 14(1999): 299-315.

Shonkwiler, J.S., and J. Englin, "Welfare losses due to livestock grazing on public lands: a count data systemwide treatment," *American Journal of Agricultural Economics* 87(2005): 302-313.

Silverman, B. "On the estimation of a probability density function by the maximum penalized likelihood method," *Annals of Statistics* 10(1982): 795–810.

Timmins, C. and J. Murdock, “A revealed preference approach to the measurement of congestion in travel cost models,” *Journal of Environmental Economics and Management* forthcoming.

Train, K., *Discrete Choice with Simulation.* Cambridge University Press: Cambridge, 2003.

Train, K., “Recreation demand models with taste variation over people,” *Land Economics*, 74:2(1998); 230-239.

von Haefen, R., “Incorporating observed choice into the construction of welfare measures from random utility models,” *Journal of Environmental Economics and Management* 45:2(2003): 145-165.

von Haefen, R., D. Phaneuf, “Identifying demand parameters in the presence of unobservables: a combined revealed and stated preference approach,” *Journal of Environmental Economics and Management* forthcoming.

Table 1 – Model Fits, In-Sample Predictions, and Compensating Surplus

Specifications				
Alternative Specific Constants?	No	Yes	No	Yes
Panel Random Parameters?	No	No	Yes	Yes
RP/SP Alberta moose hunting data from Adamowicz et al. (1997)				
Log-likelihood	-5,655.2	-5,376.7	-4,817.8	-4,521.6
Percentage improvement in log-likelihood	-	4.92%	14.8%	20.1%
Percentage absolute prediction error – all sites	30.0%	0.13%	45.6%	21.2%
CS for moose population reduction at WMU #348	-$14.11 (37.5)	-$9.47 (2.19)	-$25.00 (10.6)	-$20.91 (4.67)
Percentage prediction error at WMU #348	+10.8%	+0.12%	+16.3%	+6.60%
CS for moose population increase at WMU #344	$3.61 (2.50)	$98.34 (31.0)	$4.83 (3.19)	$73.02 (23.3)
Percentage prediction error at WMU #344	-48.4%	+0.07%	-88.1%	+26.3%
RP/SP Saskatchewan moose hunting data from Haener et al. (2001)				
Log-likelihood	-7,655.3	-7,482.3	-6,658.2	-6,547.5
Percentage improvement in log-likelihood	-	2.26%	13.0%	14.5%
Percentage absolute prediction error – all sites	26.3%	0.17%	56.8%	33.6%
CS for moose population reduction at WMZ #59	-$18.55 (7.52)	-$14.69 (2.99)	-$81.47 (11.5)	-$61.62 (9.77)
Percentage prediction error at WMZ #59	+31.5%	-0.02%	+82.9%	+30.7%
CS for moose population increase at WMZ #66	$27.54 (4.10)	$150.50 (36.9)	$22.30 (3.36)	$74.59 (14.3)
Percentage prediction error at WMZ #66	-39.2%	+0.20%	-31.3%	+13.2%
RP Mid-Atlantic beach data from Parsons et al. (1999)				
Log-likelihood	-13,160.2	-12,981.8	-11,015.8	-10,869.2
Percentage improvement in log-likelihood	-	1.36%	16.3%	17.4%
Percentage absolute	13.3%	<0.01%	27.0%	31.4%

prediction error – all sites				
CS for lost beach width at DE/MD/VA beaches	-\$6.44 (1.16)	-\$4.89 (4.62)	-\$5.64 (1.41)	-\$7.57 (3.26)
Percentage prediction error at DE/MD/VA beaches	-2.22%	<0.01%	-12.1%	-1.75%
CS for northern DE beach closings	-\$19.56 (0.64)	-\$21.88 (3.94)	-\$14.97 (1.45)	-\$16.83 (3.24)
Percentage prediction error at northern DE beaches	-5.04%	<0.01%	-6.01%	-8.28%

Robust standard errors in parentheses. All welfare estimates are per trip.

Table 2 – Alternative Strategies

Specifications								
Alternative Specific Constants?	No	Yes	No	No	Yes	Yes	Yes	Yes
Berry Contraction Mapping?	-	No	-	-	No	No	Yes	Yes
Random Parameters?	No	No	Panel	Non-Panel	Non-Panel	Panel	Non-Panel	Panel
RP/SP Alberta moose hunting data from Adamowicz et al. (1997)								
Log-likelihood	-5,655.2	-5,376.7	-4,817.8	-5,626.7	-5,368.3	-4,521.6		
Percentage absolute prediction error – all sites	30.0%	0.13%	45.6%	32.3%	0.14%	21.2%		
Unconditional CS for moose population reduction at WMU #348	-$14.11 (37.5)	-$9.47 (2.19)	-$25.00 (10.6)	-$23.09 (6.08)	-$9.56 (3.25)	-$20.91 (4.67)		
Conditional CS for moose population reduction at WMU #348	-$13.29 (2.86)	-$9.90 (1.54)	-$22.04 (3.91)	-$36.53 (7.85)	-$10.21 (7.85)	-17.01 (1.88)		
Percent. predict. error at WMU #348	+10.8%	+0.12%	+16.3%	+25.4%	+0.04%	+6.60%		
Unconditional CS for moose population increase at WMU #344	$3.61 (2.50)	$98.34 (31.0)	$4.83 (3.19)	$2.76 (1.70)	86.81 (27.4)	$73.02 (23.3)		
Conditional CS for moose population increase at WMU #344	$5.97 (7.17)	$98.00 (29.8)	$2.99 (2.37)	-$1.70 (1.62)	85.33 (23.5)	61.02 (20.6)		
Percent. predict. error at WMU #344	-48.4%	+0.07%	-88.1%	-45.2%	-0.20%	+26.3%		
RP/SP Saskatchewan moose hunting data from Haener et al. (2001)								
Log-likelihood	-7,655.3	-7,482.3	-6,658.2	-7,587.2	-7,472.9	-6,547.5		
Percentage absolute prediction error – all sites	26.3%	0.17%	56.8%	32.5%	4.84%	14.47%		
Unconditional CS for moose population reduction at WMZ #59	-$18.55 (7.52)	-$14.69 (2.99)	-$81.47 (11.5)	-$41.39 (16.3)	-$27.53 (12.1)	-$61.62 (9.77)		
Conditional CS for moose population reduction at WMZ #59	-$12.64 (5.31)	-$14.68 (2.65)	-$40.23 (4.57)	-$34.71 (12.4)	-$30.67 (12.1)	-$42.72 (4.58)		

Percent. predict. error at WMZ #59	+31.5%	-0.02%	+82.9%	+37.0%	+2.34%	+30.7%		
Unconditional CS for moose population increase at WMZ #66	$27.54 (4.10)	$150.50 (36.9)	$22.30 (3.36)	$25.14 (8.41)	$85.27 (19.0)	$74.59 (14.3)		
Conditional CS for moose population increase at WMZ #66	$32.82 (4.18)	$150.62 (36.3)	$26.57 (3.68)	$24.50 (8.85)	$79.56 (17.0)	$80.27 (14.7)		
Percent. predict. error at WMZ #66	-39.2%	+0.20%	-31.3%	-39.6%	+3.22%	+13.2%		
RP Mid-Atlantic beach data from Parsons et al. (1999)								
Log-likelihood	-13,160.2	-12,981.8	-11,015.8	-13,021.4	-12,856.5	-10,869.2	-12,874.5	-10,962.9
Percentage absolute prediction error – all sites	13.3%	<0.01%	21.8%	11.0%	+1.81%	+25.9%	<0.01%	<0.01%
Unconditional CS for lost beach width at DE/MD/VA beaches	-$6.44 (1.16)	-$4.89 (4.62)	-$5.64 (1.41)	-$3.34 (0.60)	$7.51 (7.72)	-$7.57 (3.26)	$1.83 (3.96)	-$11.76 (4.27)
Conditional CS for lost beach width at DE/MD/VA beaches	-$6.58 (1.18)	-$4.89 (4.27)	-$7.15 (1.11)	-$3.59 (0.70)	$6.95 (7.16)	-$7.35 (1.79)	$1.78 (3.70)	-$11.53 (1.95)
Percent. predict. error at DE/MD/VA beaches	-2.22%	<0.01%	-11.8%	-2.36%	+0.35%	-2.50%	<0.01%	<0.01%
Unconditional CS for northern DE beach closings	-$19.56 (0.64)	-$21.88 (3.94)	-$14.97 (1.45)	-$12.27 (0.58)	-$11.98 (4.55)	-$16.83 (3.24)	-$11.92 (2.46)	-$22.23 (3.84)
Conditional CS for northern DE beach closings	-$20.75 (0.69)	-$22.04 (2.34)	-$16.58 (2.09)	-$13.54 (0.69)	-$13.51 (1.76)	-$19.34 (1.88)	-$13.27 (1.09)	-$23.69 (2.54)
Percent. predict. error at northern DE beaches	-5.04%	<0.01%	-6.01%	-1.20%	-0.49%	-8.47%	<0.01%	<0.01%

Robust standard errors in parentheses. All welfare estimates are per trip.

An Alternative Futures Analysis for the Little Kanawha River Watershed in West Virginia

Vishakha Maskey[25],

Michael P. Strager[26],

Charles B. Yuill[27]

Paper presented at the W1133 Annual Meetings in Big Island of Hawaii, February 17-20, 2008.

Support for this analysis was provided by the Regional Research Institute, West Virginia University. This research project is funded as a Seed Grant titled "West Virginia/ Central Appalachian Land Use Status and Trends Project".

[25] Post Doctoral Fellow, Regional Research Institute, West Virginia University
[26] Assistant Professor, Natural Resource Economics, West Virginia University
[27] Associate Professor, Landscape Architecture, West Virginia University

An Alternative Futures Analysis for the Little Kanawha River Watershed in West Virginia

Abstract

The Little Kanawha River watershed in West Virginia has been identified to be one of fifteen watersheds in the United States that is projected to have the greatest amount of land conversion during the period from 2002 to 2030 (Steinitz et al., 2005). The contributing factors for land conversion in the region are resource extraction and a disproportional population growth due to suburban sprawl. The outcome of land use change has an impact on water quality, terrestrial and aquatic habitat, and biodiversity. To measure and analyze this change, an alternative futures analysis was used in this study to model scenarios with Geographic Information Systems (GIS) and help guide future policies that are sustainable by balancing the environment and development in the Little Kanawha River Watershed. The major objective of this analysis was to map the likely dispersion or future growth patterns and the impact on water quality and biodiversity. The results indicate that water quality (total suspended solids) is impacted from current development patterns as is the biodiversity (bird community index) and will continue to degrade unless an integrated planning approach which considers preservation of large intact forested lands is implemented.

Keywords: alternative futures analysis, sustainability, geographic information systems, Little Kanawha River, biodiversity, hydrology

1. Introduction

Human land use practices have influenced natural resources at local, regional and global scales (Turner, 1990). Landscapes in West Virginia have been subject to a wide range of land uses, including resource extraction. Various regions in the state are facing land use change and development pressure, such as the Little Kanawha watershed, the Eastern Panhandle, and the North Branch Watershed of the Potomac River. Communities are facing concerns regarding surface and ground water quality from animal waste, surface mines, and stream sedimentation from mining, development, and forestry practices. In addition, sub-urban sprawl from new residential and second home development are also issues related to land use change. These land use activities are contributing to environmental problems.

The evaluation of land use and cover is an extremely important activity for current land management (Kepner et al., 2004). Assessing different land use scenarios using sensitivity analysis provides a foundation for informed decision-making. This study applies such a methodology for the Little Kanawha River watershed in West Virginia. The watershed is projected to be one of fifteen in the entire United States that will have the greatest amount of land conversion from 2002 to 2030 with over 225,000 acres of land projected to be modified (Steinitz et al., 2005). An alternative futures analysis framework was applied to evaluate the different land use predictions and outcomes on water quality and biodiversity. Alternative futures analysis balances environmental and economic aspects in planning and is extremely applicable to changing regions such as the Little Kanawha River Watershed. As part of the approach several policy choices can be considered using sensitivity analysis. The goal of this study is to develop future land use growth predictions utilizing past and expected future population and growth projections for the Little Kanawha watershed.

The analysis is divided into three scenarios of 1) environmental resource protection scenario, 2) unregulated growth scenario and 3) balanced environmental protection and growth scenario. We used historical spatial and temporal data models for future land use changes for the future with a land use change GIS model. The outcomes are evaluated with hydrology and biodiversity measurements.

2. Literature Review

Examples of scenario based alternative futures analysis include those conducted in Monroe county, Pennsylvania; the region of Camp Pendleton, California; the Willamette River Basin in Western Oregon; the Southern Rocky Mountains in Alberta; the California Mojave Desert; and the Iowa Corn Belt (Steinitz et al., 2003). Additionally, Blackberry creek Watershed in Kane County and Chico Watershed in Kitsap County, Illinois are a few case studies that use alternative futures analysis for community-based decision making in environmental planning.

The Blackberry Creek watershed resource planning committee with the assistance of numerous municipal, county, regional, state, and federal agencies, as well as private consultants, developed the Blackberry Creek Watershed Management Plan (Environmental Law Institute, 2004). Conservation and conventional versions of the template were developed on a hypothetical 40 acre parcel for a range of land uses including commercial, residential, and agriculture, as well as wetlands and streams. A conservation template was created based on preserving natural hydrologic mechanisms, minimizing changes in hydrology and water quality caused by land development. A conventional template was based on the current practice for site design and stormwater management, that collects, conveys and detains storm water rather than distributes, infiltrates and retains. The plan was adopted by most of the municipalities mostly because flooding was the major problem that impacted most of the areas in the watershed. The focus of the project was to protect the streams and wetlands from direct modifications, and to prevent degradation of watershed hydrology, and water quality of streams and wetlands.

The Willamette River Basin is a region facing population growth creating land use and water use concerns, habitat loss, and loss of forest land (USEPA, 2002). Their alternative futures analysis had three scenarios of conservation, plan trend and development. Conservation trend placed emphasis on ecosystem protection and restoration. Planned trend represented future landscapes if current polices are implemented and recent trends continues. Development trend loosens current policies to allow free market forces across all components of the landscape. A basic assumption was made that population increases were to grow in a similar ratio that was linear to the year 2050. The three alternative scenarios were compared to present day, historical landscapes and future endpoints in terms of terrestrial wildlife, water availability, small streams and the Willamette River. The results from the analysis were discussed by stake holder groups in developing visions for the area's future and restoration strategy. In addition, the results influenced future decisions of resource use.

Kepner et al. (2004) used Automated Geospatial Watershed Assessment (AGWA) for the San Pedro River Basin alternative futures analysis. A base time was year 2000 projected to 2020. Three scenarios of constrained, planned, and open were evaluated. In addition, the Soil Water Assessment Tool (SWAT) was used for simulation of the large watershed. The results showed significant alteration on hydrologic responses in the watershed due to urbanization and land use practices. The increase in amount of run-off, sediment discharge and loss of surface water access to the ground water table were predicted from the simulation.

The Chico Watershed alternative futures analysis was used to guide community planning and natural resource protection (Parametrix Inc., 2003). The Chico Creek Watershed was utilized as a pilot project for alternative futures planning due to its healthy salmon runs, large tracts of forestland, two large lakes, and the increasing demands of development within its boundaries. Four different scenarios analyzed the effect on water quality, water quantity and fish and wildlife habitat. The analysis also assessed the potential benefits and impacts of those future land use scenarios. The result was a developed watershed plan based on natural resource protection with citizen's involvement.

Alternative futures with three scenarios of production, water quality and biodiversity were utilized in the Iowa Corn Belt, in response to environmental degradation resulting from agricultural practices (Santelman et al., 2004). Their spatial model evaluated the farmland management policies and the results evaluated impacts of land use change on water quality, social and economic goals, and native biodiversity.

The Soil Water Assessment Tool (SWAT) was used in alternative futures scenarios for two watersheds (Buck and Walnut Creek) in Iowa (Vaché et al., 2002). The model focused on water discharge, annual sediment loads, and nitrate in watersheds based on different agricultural management practices. Three scenarios were developed, namely, current trends in agricultural practices scenario, water quality concerned scenario, and biodiversity protection and restoration.

The Ecosystem Landscape Modeling System (ELMS) was used for evaluating the potential ecological and economic impacts of future landscape changes in areas of the Rocky Mountain that is facing rapid growth (Prato, 2005). ELMS consists of an economic, land use change, ecological assessment, and policy models. The economic model used IMPLAN and various assumptions to estimate changes in employment and output for alternative future growth rates for sectors in the study area. Future changes in employment and output were translated into land use requirements for residential housing and commercial establishments. The ecological assessment model evaluated impacts of land use changes on potential and realized habitat for selected species. The policy model specified alternative residential and commercial development, infrastructure expansion, and natural resource conservation policies that were incorporated in the economic and land use change models (Prato, 2005). The result evaluated tradeoffs between alterative scenarios associated with future growth and development.

3. Study Area: Little Kanawha River Watershed

The Little Kanawha River watershed is located in central West Virginia with an approximate area of 2,307 square miles (USEPA, 2000) (Figure 1). Forested lands and agriculture constitutes 77% and 16% of the land cover within the watershed. The rest of the watershed is compromised of water, urban built-up land, and transportation usage (USEPA, 2000).

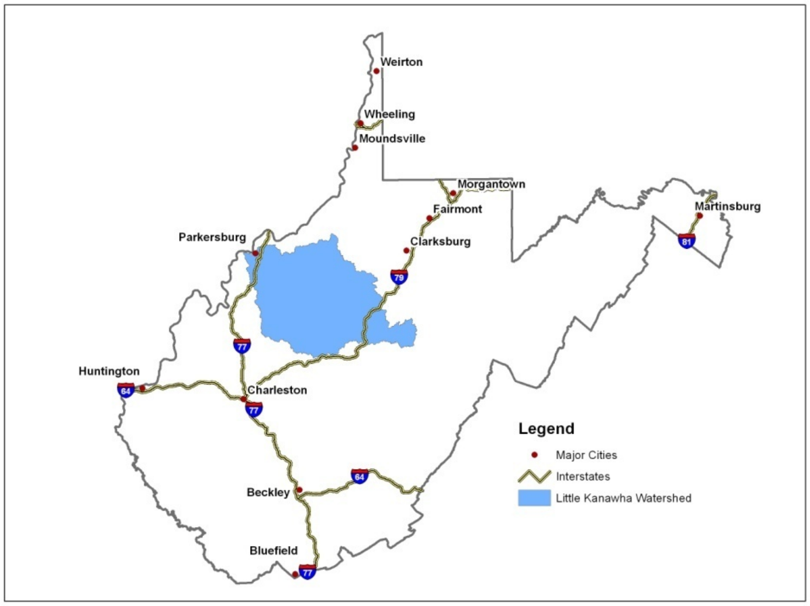


Figure 1. Location of the Little Kanawha River Basin, West Virginia.

The watershed intersects sixteen West Virginia counties and there are four reported watershed protection groups involved in Little Kanawha River. Gilmer watershed coalition is involved in water quantity problems (Flood control, flood warning system), water quality testing, trash and debris removal, restoration work and working with agencies. Friends of the Little Kanawha, Cedarville Community Association and Huges Creek watershed associations serve as alliance and council for watershed adoption (EPA, 2007).

Substantial land use changes have occurred and continue to occur in the Little Kanawha. There has been significant change in agriculture, forestry, mining, accessibility and infrastructure and settlement patterns during the past fifty years (Figure 2). Major land use conversion due to a decrease of agriculture, reforestation, development of a chemical industry and suburban sprawl are the driving factors of land use change.

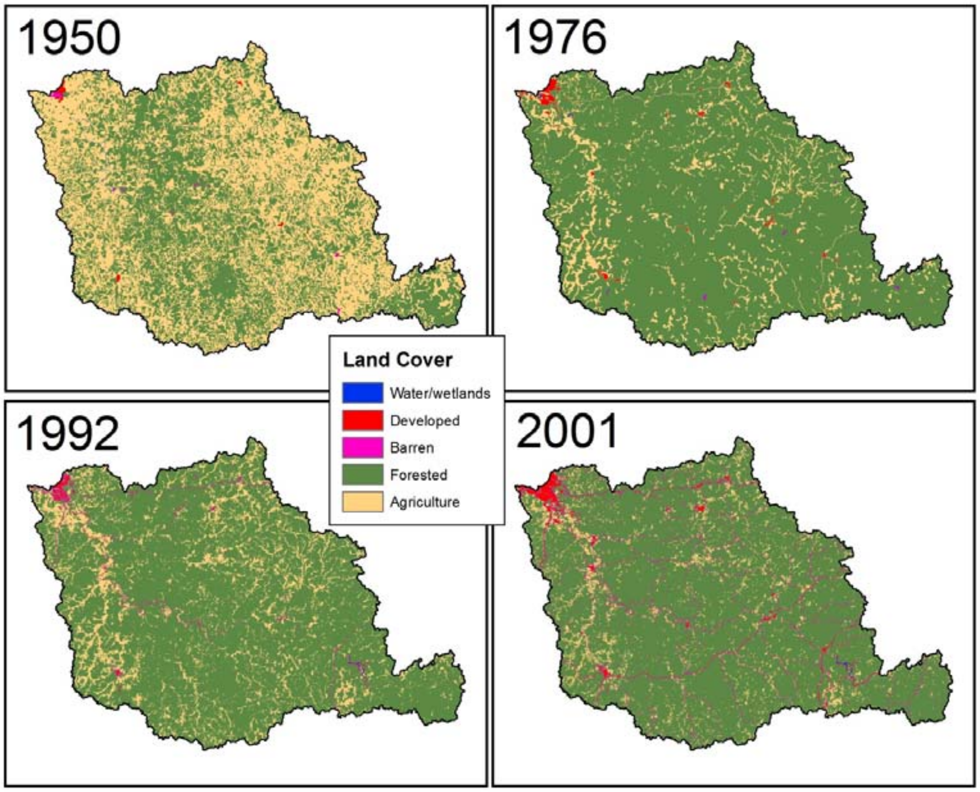


Figure 2. Historical Land use and cover change in the Little Kanawha River Basin, West Virginia.

Due to many contiguous forested areas, the Little Kanawha River watershed is high in species richness (Strager and Yuill, 2002, Figure 3) in many of the headwater locations. Maintaining biodiversity is a significant aspect of the watershed. Land conversion and forest fragmentation due to human activities pose a threat to habitat and biodiversity (Collinge, 1996). Concerns regarding biodiversity protection and habitat restoration are important to this region.

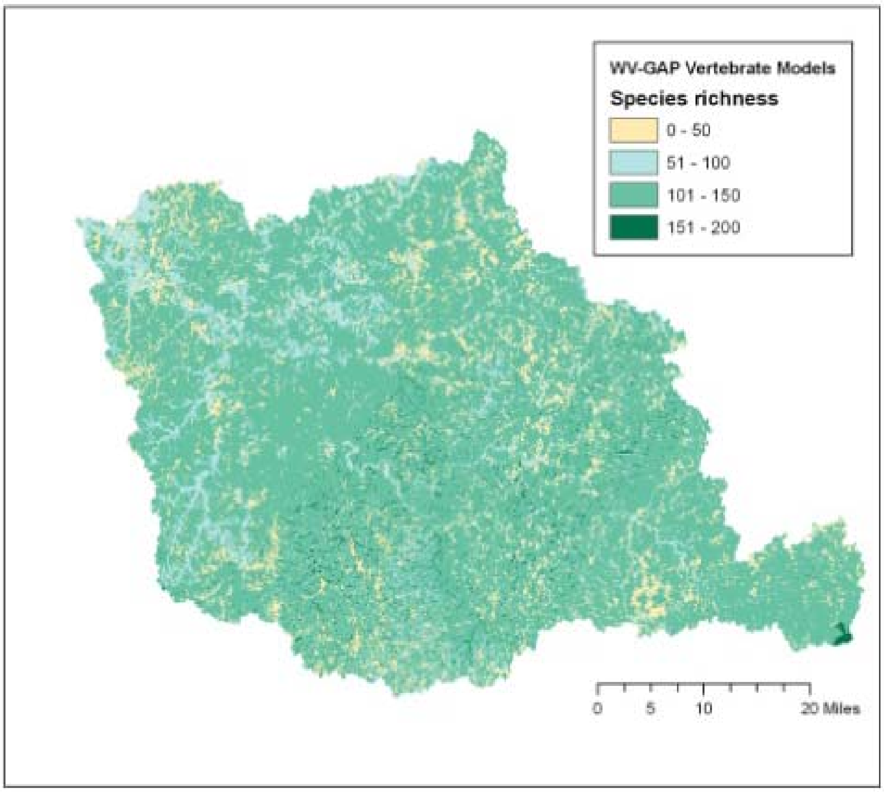


Figure 3. Species distribution in the Little Kanawha River Basin, West Virginia.

In addition to biodiversity, water quality and flooding are additional concerns in the watershed. Flooding was one of the major problems in Little Kanawha River around 1970-1989 (Smith 2008) and continues mainly due to the impact of altering natural vegetation on steep slopes. The alterations also impact water quality as many of the streams are listed as impaired or not meeting their designated uses as shown in Figure 4. The streams tend to be located in both the headwater and mainstem of the river channel indicating diverse issues with point and nonpoint source pollution. According to the USEPA 303 (d) list of impaired waters, there are 47 river miles of the Little Kanawha River from the Burnsville Dam to its convergence with the Ohio River that are not meeting their designated uses of either cold or warm water fisheries, or water contact recreation.

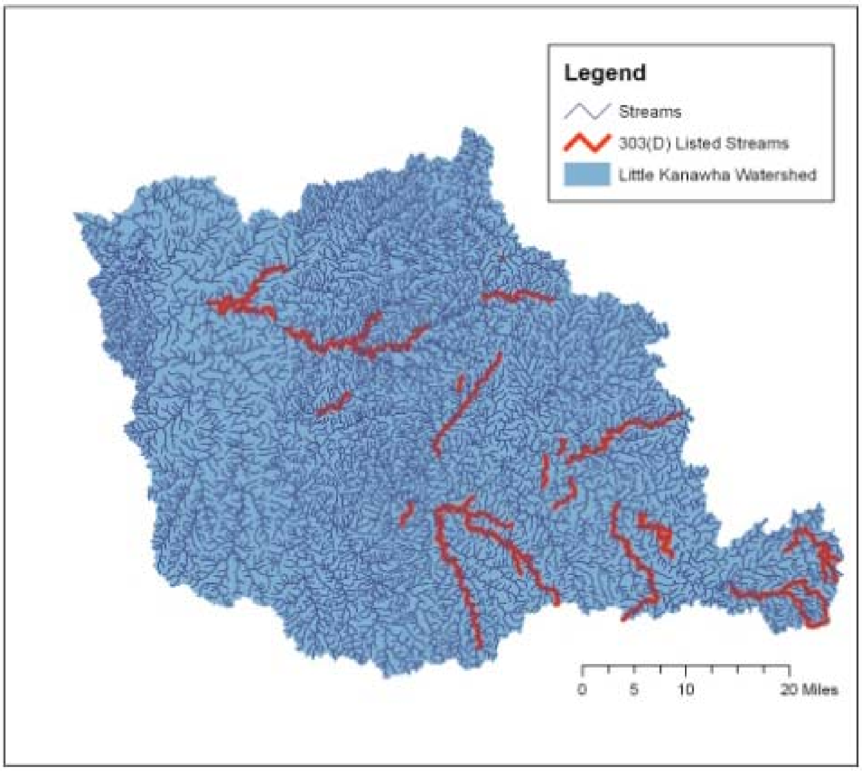


Figure 4. EPA listing of impaired streams for the Little Kanawha River Watershed

Daniels (1993) defines rural sprawl as low-density residential development scattered outside of suburbs and cities, and as commercial strip development along roads outside cities. Rural sprawl with rural residential development at exurban densities according to Theobald (2003) is defined as areas typically 1.7 to 20 acres per housing unit for exurban sprawl. In some states, exurban areas are defined as having between 1.7 and 40 acres per housing unit, depending on state land use laws. Rural areas have >20 acres per housing unit (or >40 acres). Issues faced in the Little Kanawha watershed are an increase in developed land from 0.8% to 6%, also a decrease in agricultural land from 12% to 7% within 1992-2001. More specifically, there is a disproportional amount of development which has impacts on watershed ecology, as well as biodiversity. In addition, a decrease is projected in forest land due to increased exurban sprawl (Steinitz et al., 2005).

Sustainable development entails maintaining development and conservation, which includes protecting water quality, habitat, biodiversity, and floodplain areas within the context of human habitation and continued sprawl (Theobald, 2003).

4. Data and Methods

Following the approach in the Willamette River by USEPA (2002), we examined policies for future land use and cover scenarios consisting of an environmental resource protection scenario, unregulated growth scenario, and a balanced environmental protection and growth scenario. A general conceptual framework for the alternative futures analysis process is shown in Figure 5. The environmental resource protection scenario (environment) limited development to slopes less than 5%, protected riparian areas that were within 100 meters of streams, avoided development on hydric soils or existing mapped wetlands, and maintained forested areas that had at least 200 acres of core area. The unregulated growth scenario (unconstrained) did not have any of these constraints, and the balanced environmental and growth scenario (balanced) allowed development to occur on any slopes and to fragment forests but did protect riparian areas, and wetlands.

One of the limitations of this study was an accurate measurement of economic impacts from these different scenarios on residential and commercial development that was constrained from these policies. Our initial approach was to examine property sales prices but we did not have digital parcels for this watershed.

The future land use change grids were created with these three policies in mind (environment, unconstrained, and balanced) as input in the Land Use Change Modeler ArcGIS 9.2 extension (Clark Labs, 2007). Output grids from the different scenarios were the major inputs into the Watershed Characterization and Modeling System (WCMS, 2004), and the Bird

Community Index (Jones et al., 1997). The WCMS is an ArcGIS extension that was designed to compare and illustrate loadings from nitrogen, phosphorous and total suspended solids on receiving water bodies as a result of land use and cover changes. It has a stream flow model which accounts for average annual conditions that are calibrated to USGS gauges. The Bird Community Index was developed by researchers at Penn State and EPA to determine habitat requirements at the landscape level for neotropical and migratory species birds. It is an overall index of landscape condition in which a higher score for neotropical species is more unique and acknowledged as better for overall landscape quality.

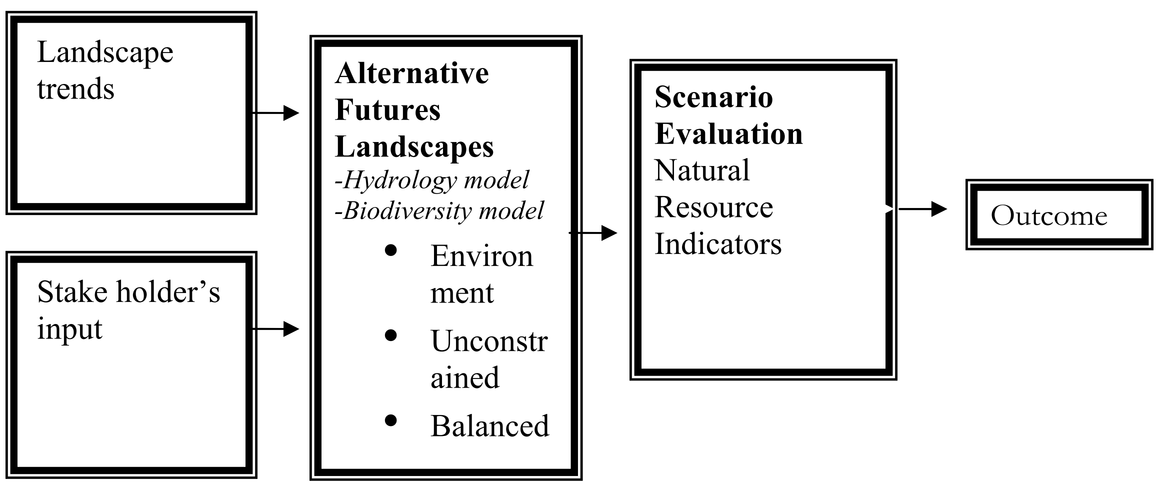


Figure 5. Alternative futures process applied in the Little Kanawha River Basin, West Virginia.

5. Results

The results of the land use and cover modeling gave insights into expected conversion of land under the different environment, unconstrained and balanced conditions. The modeling was primarily driven by existing infrastructure such as roads and already developed areas. Some of the rural areas were difficult to model since many of the traditional drivers which may spur development did not exist in the rural areas (i.e areas that may have favorable property values or parcel sizes for development). Despite this limitation, the model output was run in both the WCMS water quality model and the Bird Community Index as proxies for natural resource indicators. The results are presented as output in table 1. It should be noted that these evaluations are done for the entire watershed boundary and extent which may hide or detract from local effects.

Table 1. Results for Total Suspended Solids and Bird Community Index

Scenario	**Total Suspended Solids (cumulative mainstem tributary loadings to the Little Kanawha)**	**Bird Community Index (measured in total landscape area of neotropical versus migratory species)**
Environment	46% less in Kg/Yr compared to current base levels	34% area improvement compared to current levels
Unconstrained	14% more in Kg/Yr compared to current base levels	4% area improvement compared to current levels
Balanced	9% more in Kg/Yr compared to current base levels	18% area improvement compared to current levels

6. Discussion

It was interesting to note that the water quality difference from choosing an environmental development scenario greatly benefits water quality over the other options. This is believed to be a result of the terrain being constrained to the riparian area because of steep and rugged slopes. While we did not have a flood potential indicator, we feel that it would have also shown significant improvement with the environmental scenario based on this reason. The improvements to the bird community index were also noted with the environmental scenario which resulted in an increase of the area of neotropical of 34%. One of the main drivers of this result could be attributed to the environment scenario protected large intact forest areas from disturbances such as clear cut forest management and surface mining. Areas with a low index contain more non forest land uses due to mining, agriculture, mining, timbering, urban/residential development. These areas are fine for generalist species such as European Starlings, American Crows, and Blue Jays. Areas with a high index are primary forested and provide habitat for many neotropical birds including Cerulean Warbler, Scarlet Tanager, and Louisiana Waterthrush.

The unconstrained option resulted in a much lower area for neotropical birds as part of the bird community index and an increase in total suspended solids of 14% compared to base levels. It seems as if the development patterns at the landscape level clearly are due to resource extraction and conversion from forests to other more barren classes. One of the policy recommendations could be to focus any conservation area acquisition on the large intact forested areas since the Jones et al. (1997) reported this as a major area for conservation goals as well.

The balanced scenario was somewhat centered between the extremes of the previous mentioned results however this is difficult to compare since we did not yet have economic development benefits to include in the analysis. For example, much of the economic gains and tax base generated from development was not accounted for in this approach. Future work will attempt to identify more economic benefits as indicators across these landscape and policy options.

Even though our results are limited at this time, we plan to continue development of additional indicators that relate more to human use and economic benefits. We feel at that time our results can begin to assist managers and planners in making informed decisions related to future land use practices that can be used by local policy makers, development planners, stakeholders, and communities in choosing sustainable land use management plans. The projections could provide a technical statement for future valuation projects.

7. Reference

Clark Labs. 2007. The Land Change Modeler for ArcGIS. Software Extension. Clark Labs, Clark University, Worchester, MA.

Collinge, S. K. 1996. Ecological consequences of habitat fragmentation: implications for landscape architecture and planning, *Landscape and Urban Planning,* Vol. 36 (1): 59-77.

Danielson, B.J., R.C. Corry, M.E. Clark, S. Polasky, R.M. Cruse, J. Sifneos, H. Rustigian, C. Coiner, J.Wu and D. Debinski. 2004. Assessing alternative futures for agriculture in Iowa, U.S.A. Landscape Ecology 19: 357–374.

Environmental Law Institute. 2004. Alternative Futures Analysis-Blackberry Creek Watershed, Illinois. Watershed Resource Consultants and the Conservation Foundation. Kane County Department of Environmental Management.

Hulse, D., J. Eilers, K. Freemark, C. Hummon, and D. White, 2000. Planning Alternative Future Landscapes in Oregon: Evaluating Effects on Water Quality and Biodiversity. Landscape Journal 19 (2):1-19.

Jones, B.K, K.H. Ritters, J. D. Wickham, R.D. Tankersley, R.V. ONeill, D.J. Chaloud, E. R. Smith, and A.C. Neale. 1997 An Ecological Assessment of United States Mid-Atlantic Region: A Landscape Atlas.

Kepner, W.G., D.J. Semmens, S.D. Bassett, D.A. Mouat, and D.C. Goodrich. 2004.

“Scenario Analysis for the San Pedro River, Analyzing Hydrological Consequences of a Future Environment.” Journal of Environmental Monitoring and Assessment, 94: 115-127. Kluwer Academic Publishers (http://www.epa.gov/nerlesd1/land-sci/pdf/scenario_spedro.pdf).

Parametrix Inc. 2003. Kitnap County Chico Watershed Alternative Analysis, Process and Recommendations. Department of Community Development, Kitnap County.

Prato, T. 2005. Modeling ecological impacts of landscape change. Environmental Modelling & Software, Vol. 20 (10): 1359-1363.

Santelmann, M.V., D. White, K. Freemark, J.I. Nassauer, J.M. Eilers, K.B. Vaché, Shearer, A. W., D. A. Mouat, S. D. Bassett, M. W. Binford, C. W. Johnson and J. A. Saarinen. 2006. Examining development-related uncertainties for environmental management: Strategic planning scenarios in Southern California. Landscape and Urban Planning, Vol. 77 (4): 359-381.

Steinitz, C., Arias, H., Bassett, S., Flaxman, M., Goode, T., Maddock T. III, Mouat, D., Peiser, R. and Shearer, A.: 2003, Alternative Futures for Changing Landscapes. The Upper San Pedro River Basin in Arizona and Sonora, Island Press, Washington, DC, USA.

Strager, J. M. and C. B. Yuill. 2002. WV Gap Analysis Project – Final Report. Natural Resource Analysis Center, West Virginia University, Morgantown, WV. Available online at www.nrac.wvu.edu

Smith, Deana. 2008. Friends of Little Kanawha. Telephone Communication on March 13.

Theobald, D. M. 2003. Defining and mapping rural sprawl. Examples from the Northwest U.S. Natural Resource Ecology Lab and Dept. of Recreation & Tourism Colorado State University, Fort Collins, CO 80523.

Turner, B.L. (ed.): 1990. The Earth as Transformed by Human Action, Cambridge University Press with Clark University, Cambridge, MA, USA.

USEPA. 2002. Willamette Basin Alternative Futures Analysis, Office of Research and Development.

USEPA. 2000. Metals TMDL for Little Kanawha River Watershed, West Virginia. Final Report, US Environmental Protection Agency, Region 3, 1650 Arch Street, Philadelphia, PA

USGS. 2005. Summary of Significant Floods in the United States, Puerto Rico, and the Virgin Islands, 1970 through 1989, Water-Supply Paper 2502. U. S. Department of Interior, U.S. Geological Survey.

Vaché, K. B., J. M. Eilers, and M. V. Santelmann. 2002. Water Quality Modeling of Alternative Agricultural Scenarios in the U.S. Corn Belt. Journal of the American Water Resources Association, Vol. 38, No. 3 pp 773-787.

WCMS. 2004. The Watershed Characterization and Modeling System. Version 9.2. Natural Resource Analysis Center, Morgantown, WV. Available online at www.nrac.wvu.edu.

A discrete-choice model of annual license demand

Eric English

Stratus Consulting Inc., Boulder, CO

W2133: Benefits and Costs of Resource Policies Affecting Public and Private Lands

Big Island, HI, February 17-20, 2008

Abstract: A multiple-site discrete-choice model is estimated based on demand for recreational licenses rather than demand for recreational trips. Choices depend on annual license fees, trip cost and site characteristics, which act through their influence on annual consumer surplus. The model was developed to examine recreational shellfishing in southeastern Massachusetts, where a conventional trip-based model is complicated by the presence of license fees that differ among local jurisdictions. The analysis illustrates the value of exploiting information on entry and exit of participants when modeling recreation behavior.

A discrete-choice model of annual license demand

Introduction

In most recreation demand models the source of information about preferences is data on recreation trips. Demand for trips is regressed on trip prices and site characteristics to obtain estimates of preference parameters. A problem may arise if consumer choices are influenced by annual fees in addition to per-trip prices. For example, coastal towns in southeastern Massachusetts each sell a separate recreational license for access to local shellfish beds and annual license fees vary across towns by $100 or more. A researcher may wish to examine shellfishing demand and substitution throughout the region, but a multiple-site model of recreation trips would falsely ascribe the reduced demand in high-fee towns to local resource characteristics. Methods available in the literature do not appear to address this type of situation.

Motivated by this problem we develop a model of recreational shellfishing using observed demand for annual shellfish licenses. Preference orderings are based on expected consumer surplus for a license and individuals choose the highest-valued alternative. Expected consumer surplus is a function of expected trip demand, which depends on the prices and characteristics of site visits. The model accounts for the influence of license fees because consumer surplus is calculated net of the fees.

The analysis may be of interest in several respects. First, we show that entry and exit behavior, such as the decision whether to purchase an annual license, contains information about the value of resource use that can be exploited in discrete-choice analysis of recreation. Only in special circumstances, however, will license-demand data alone be sufficient to identify a valuation model. Second, our results suggest that certain types of participation-level decisions may be addressed with greater precision even in trip-based models of recreation demand. Annual fishing licenses, season passes at a beach, or slips at a marina all have value and are acquired at a price, but the resulting impacts on behavior are often absorbed into the statistical noise of conventional trip-demand models. Finally, this study appears to offer the first evidence available in the literature on the value of recreational shellfishing.

License Demand and Recreation Participation

Previous studies of license demand have not modeled individual-level choices, but have instead obtained estimates of consumer surplus from the area under a market-level demand function. Loomis, Pierce, and Manfredo (2000) estimated demand for deer and elk hunting licenses in Colorado; Sun, van Kooten, and Voss (2005) estimated demand for wildlife hunting licenses in British Columbia; and Bennear, Stavins, and Wagner (2005) predicted demand for fishing licenses by state throughout the U.S. All three articles used time-series data to determine the response in license demand to changes in inflation-adjusted license fees.

Estimating license demand is closely related to modeling recreation participation. The choice to participate during any given year is equivalent to the choice to purchase a license. A substantial literature has developed regarding feasible methods to predict the number of participants in a population. Smith and Desvousges (1985), Shaw (1988) and Bockstael *et al*. (1990) were among the first to examine the issue by dividing a statistical distribution of demand outcomes into the zero and positive range. The size of the participant group was determined by the probability mass on positive demand outcomes in a normal (Tobit) or Poisson distribution. Extensions to this approach included negative binomial models (Grogger and Carson 1991; Haab and McConnell 1996; Gillig, Ozuna, and Griffin 2000) and hurdle models (Gurmu and Trivedi 1996; Haab and McConnell 1996; von Haefen, Massey, and Adamowicz 2005). These and other developments improved predictions of the size of the participant group but did not specifically address the choice to participate or the impact of license fees on recreation behavior.

Specifications for participation behavior more closely related to individual choice have been developed in two types of models. The Kuhn-Tucker model (Phaneuf 1999) is derived directly from individual utility functions. Draws from a distribution of preferences generate utility-rankings over all possible recreation-consumption bundles, including the choice not to participate. License demand could be incorporated into the Kuhn-Tucker model by adding to the available bundles access to a licensed activity and payment of fees, though previous applications have not done so (*e.g.*, von Haefen and Phaneuf 2003). (Forthcoming) derived a logit-based specification for the choice to participate using individual demand functions. In (forthcoming)'s "choke price" model, heterogeneity in preferences corresponds to a distribution of individual demand functions and the threshold for participation corresponds to an individual's choke price. The effect of license fees on individual choice is captured by change in the choke price. The model developed in the next section uses an approach similar to the specification in (forthcoming) to predict participation choices in the presence of site-specific license fees.

Model

The model is derived in the context of random utility maximization (McFadden 1974). Utility rankings $\{U_{nj}\}$ describe license-purchase decisions, where n denotes individuals in the population and j denotes alternatives $j \in J$. There are $J-1$ sites where a license can be purchased and the choice not to purchase a license is denoted by $j = 0$. With nonparticipation captured in U_{n0}, the choice alternatives are exhaustive. It is also assumed that alternatives are mutually exclusive. For the shellfishing data analyzed below, we believe this assumption is behaviorally insignificant because no individuals were identified who purchased a license in more than one location.[28] Without loss of generality let $U_{n0} = 0$ and let utility for a license at each of the sites be described by

$$U_{n,j>0} = g\left(p_{nj}, \eta_{nj}, fee_{nj}\right). \tag{1}$$

For individual n and site j utility depends on trip prices p_{nj}, preferences for site characteristics η_{nj} and annual license fees fee_{nj}.

Utility for a license should depend on the value of expected trips, which can be incorporated into the functional form for (1). For each site we define CS_{nj} as the area under an individual's expected demand function for trips conditional on access to site j, less the license fee for site j. The expected demand function gives expected trips at the start of the season as a function of trip price, which is the cost of travel to a site. For nonparticipation we define CS_{n0} =

[28] Of the 11 shellfishing jurisdictions providing data on license purchases, five provided the names of license holders and no duplicate names were identified.

0. Following simplifications applied in previous studies[29] we ignore within-season discounting of the value of trips, uncertainty regarding expected trip demand and income effects associated with changes in the price of trips. The effects of within-season discounting may lead to a discrepancy between the value of money spent on a license and the value of trips taken later in the season. Uncertainty in trip demand may lead to further discounting of future trips. We believe the extent of these effects would be difficult to identify separately from the consumer surplus area associated with expected demand, the value to which any adjustments would be applied. Income effects are likely to be modest because the cost of local recreation trips represents a small share of income for most people. While further examination of these effects might be possible using data on actual trips, any assumption about the relationship between actual trips and expected trips would entail difficulties of its own. The difficulties would include potential discrepancies between expected and actual trips (Huang, Haab, and Whitehead 1997) and the likely outcome of zero trips for some license holders (*e.g.*, forthcoming). Recognizing the potential uncertainties, we adopt the above assumptions and let $U_{nj} = CS_{nj}$. When CS_{nj} is positive, it is equivalent to the net monetary value of a license at the time of purchase. Since $CS_{n0} = 0$ is always available, a correspondence between $\max_j\{CS_{nj}\}$ and $\max_j\{U_{nj}\}$ is satisfied for all observed choices.

In discrete-choice models it is common for demand functions to take a logistic form. For examples specifically related to annual recreation demand, see Morey (1999), Parsons, Jakus, and Tomasi (1999) and (forthcoming). Following this practice, the functional form for expected

[29] Bennear, Stavins, and Wagner (2005) also assumed that effects of uncertainty and within-season discounting were small enough to ignore. Their study appears to be the only previous attempt to directly address the connection between license value and trip value. The third assumption regarding zero income effects is common in the recreation demand literature (*e.g.*, Morey 1999; Parsons 2004).

trips conditional on the purchase of a license at site j is given by

$$q_{n,j>0} = D\frac{e^{\alpha p_{nj}+\eta_{nj}}}{e^{\alpha p_{nj}+\eta_{nj}}+1}. \tag{2}$$

The constant D sets an upper limit on an individual's annual trips. As in any discrete-choice model, η_{nj} could be further parameterized by regressing on observed site characteristics. Note that the area under (2) represented by the integral $\int q_{nj}(p)dp$ from p_{nj} to ∞ is given by $D\ln(e^{\alpha p_{nj}+\eta_{nj}}+1)/(-\alpha)$, as described by McConnell (1995).

To account for entry and exit from the activity and the effect of license fees on participation, individual demands must incorporate a finite choke price. Let $C(p_{nj}^c)$ be the area under (2) above the choke price p_{nj}^c. In the presence of a finite choke price, the area under an individual's demand function becomes $D\ln(e^{\alpha p_{nj}+\eta_{nj}}+1)/(-\alpha) - C(p_{nj}^c)$. Subtracting $C(p_{nj}^c)$ accounts for the threshold level of value that must be obtained from expected trips in order to compensate for the cost of a license and induce an individual to participate. If $C(p_{nj}^c)$ takes a value greater than the license fee, the choke price is finite even in the absence of a fee.

The choke price may be further defined based on two prior expectations. First, we expect that in the absence of a license fee demand at price p_{nj}^c would be somewhat consistent across individuals and sites. For example, without a fee most participants who are at the point of indifference between participating and not participating at a given site might have demand in the neighborhood of a single trip. We also expect that $\partial p^c / \partial \eta > 0$, since higher site utilities correspond to greater willingness to pay for trips. A form for $C(\cdot)$ and $p^c(\cdot)$ that maintains these assumptions is $C(p^c(\eta_{nj}, fee_{nj})) = c + fee_{nj}$. Adopting this form, utility for sites can be written as

$$U_{n,j>0} = CS_{nj} = D\frac{\ln\left(e^{\alpha p_{nj}+\eta_{nj}}+1\right)}{-\alpha} - c - fee_{nj}\,. \tag{3}$$

Tastes for recreation sites follow a joint distribution $f(\eta)$. License demand is described by selection probabilities for the available alternatives, including nonparticipation. Specifically,

$$P_{nk} = \int_{\eta_1=-\infty}^{\infty} \cdots \int_{\eta_J=-\infty}^{\infty} I[\max_j\{U_{nj}\} = U_{nk}]\,f(\eta)\,d\eta_1 \ldots d\eta_J\,. \tag{4}$$

For any realization of η the indicator function takes a value of one if alternative k is the highest-utility option. Integrating the indicator function over the distribution $f(\eta)$ identifies the portion of the support of $f(\eta)$ associated with choice k, which is equivalent to P_{nk}.

Note that the model depends on the price of an expected trip but does not use information on the number of expected trips, which is unknown. The source of model identification should therefore be clarified. Using (3), consider the monotonic transformation of utilities given by $U'_{n,j>0} = -\alpha CS_{nj} = D\ln\left(e^{\alpha p_{nj}+\eta_{nj}}+1\right) + \alpha c + \alpha fee_{nj}$. The change in utility due to an increase in the price of a trip is $\partial U'/\partial p = \alpha D e^{\alpha p_{nj}+\eta_{nj}}/(e^{\alpha p_{nj}+\eta_{nj}}+1) = \alpha q$. This makes sense, since the price of a trip affects the value of a license in proportion to expected trips. The model can determine $\partial U'/\partial p$ based on the influence of trip price on license demand. However, since q is unknown, the model is identified with respect to changes in price (meaning αq can be estimated) only if α is identified independently of price. Variation in fee_{nj} insures that this is the case.

Data

The shellfishing license data consist of license purchases in 2004 by town of residence for 11 shellfishing sites. The data were compiled from materials provided by state and local resource management officials. Ten of the sites are towns along the southeastern shore of Massachusetts from Scituate south, excluding Marshfield, Fairhaven and Cape Cod. The eleventh site is the state of Rhode Island. Scituate is the first location south of Boston where recreational shellfishing is allowed. Marshfield did not allow shellfishing in 2004 and data for the town of Fairhaven could not be obtained. Cape Cod was not included in the analysis because it is geographically distinct from the mainland sites. Trips to Cape Cod sites by residents of the mainland may also involve a large number of multiple-day trips which could complicate the analysis.

An overview of license purchases and fees is presented in Table 1. Sites are listed geographically from north to south. A total of 6,225 licenses were purchased for the 11 sites. Residents of a town where shellfishing is permitted pay a lower license fee than non-residents, and senior residents (over 65) pay the lowest fees. Rhode Island is included in the model to account for choices available to Massachusetts residents, but license purchases by Rhode Island residents are excluded from the data. The total region delineated for the travel-cost analysis includes all Massachusetts municipalities east of Worcester and south of Boston, excluding Cape

Cod. This amounts to 127 cities and towns, accounting for all but 22 of the 6,225 licenses in the data and including 15 towns where no one purchased a license.

[Table 1]

No survey was conducted for this analysis and demographic variables enter the model as aggregate-level data obtained from public sources. Specifically, all residents of the region are assigned demographic characteristics based on average statistics for their town of residence. Town-level data are not available from the U.S. Census and were instead obtained from the Massachusetts Institute for Social and Economic Research and a statistical guide to Massachusetts towns published by the New York Times Company.[30] Demographic variables include median household income divided by $10,000 ("income"), the percentage of adults with a college degree ("education") and the percentage of households with children under 18 ("kids"). Previous studies that have used aggregate-level variables include Shonkwiler and Englin (2005), Englin, Boxall, and Watson (1998), and Englin and Shonkwiler (1995). The advantages and drawbacks of using aggregate data are discussed in Hellerstein (1995) and Moeltner (2003).

Travel distances were measured from each town of origin to the 11 sites. Round-trip distances were converted to prices at a rate of $0.682 per mile based on marginal driving expenses estimated by the American Automobile Association and one-third average hourly earnings for Massachusetts households as reported by the U.S. Census. Estimating the time-cost

[30] The demographic data are available at www.umass.edu/miser and www.boston.com and were accessed for this study on January 21, 2007.

of driving using some variant of one-third the wage rate originated with Cesario (1976) and follows studies such as Train (1998), Parsons, Plantinga, and Boyle (2000) and Moeltner (2003).

We focus only on access value for sites so our data does not include information on site characteristics. The J - 1 means for $f(\eta)$ therefore function as alternative-specific constants. It is possible that site characteristics would provide some information about the value of sites that is not available from observed choices alone. However, most authors suggest that observed choices provide the best information on the utility of choice alternatives and that any mismatch with observed site characteristics results from the presence of unobserved characteristics (Hausman, Leonard, and McFadden 1995; Train, McFadden, and Johnson 2000; Murdock 2006). The use of alternative-specific constants is therefore well suited to estimating the value of shellfishing.

Estimation

A transformation of (1) through (4) will assist with estimation and interpretation of model parameters. Specifically, we would like to avoid an excess of covariance parameters to describe $f(\eta)$ and instead use any correlations between random terms to represent substitution patterns analogous to those of a standard multiple-site participation model (*e.g*., Morey 1999). As presented in (1) through (4) preferences $f(\eta)$ would be dominated by correlations describing demand heterogeneity rather than site substitution. For example, many individuals are not interested in participating in shellfishing and they would be represented by low draws of η_j for all $J-1$ sites. Demand heterogeneity can be distinguished from substitution-related correlations by the transformation $\eta_{nj} = \gamma_j + \varepsilon_{nj} - V_{n0}$, where both ε_{nj} and V_{n0} are random. V_{n0} is the utility of alternative activities relative to shellfishing trips. Variation in V_{n0} represents correlations across all η_j, corresponding to heterogeneity in demand for shellfishing. Substitution patterns among sites are represented by correlations in $f(\varepsilon)$. The terms γ_j are site constants, so that $f(\varepsilon)$ is centered on zero. Consumer surplus in (3) can now be rewritten as

$$U_{n,j>0} = CS_{nj} = D\frac{\ln\left(e^{\alpha p_{nj} + \gamma_j + \varepsilon_{nj}} + e^{V_{n0}}\right) - V_{n0}}{-\alpha} - c - fee_{nj}\,. \tag{5}$$

This formulation more closely resembles the "choke price" model described in (forthcoming). In discrete-choice analysis, errors associated with constant terms typically take a logistic or normal form. We use a normal distribution for $f(\varepsilon)$ and $f(V_0)$.

Following (forthcoming), the threshold value of V_{n0} associated with an individual's decision to participate at a given site is found by setting $CS_{nj} = 0$ and solving for V_0^*:[31]

---

[31] The steps in the calculation are:

$$0 = D\frac{\ln\left(e^{\alpha p_{nj} + \gamma_j + \varepsilon_{nj}} + e^{V_0^*}\right) - V_0^*}{-\alpha} - c - fee_{nj}\,,$$

$$\frac{-\alpha\left(c + fee_{nj}\right)}{D} = \ln\left(e^{\alpha p_{nj} + \gamma_j + \varepsilon_{nj}} + e^{V_0^*}\right) - V_0^*\,,$$

$$e^{-\alpha\left(c + fee_{nj}\right)/D} = \left(e^{\alpha p_{nj} + \gamma_j + \varepsilon_{nj}} + e^{V_0^*}\right)e^{-V_0^*}\,,$$

$$e^{-\alpha\left(c + fee_{nj}\right)/D} - 1 = \frac{e^{\alpha p_{nj} + \gamma_j + \varepsilon_{nj}}}{e^{V_0^*}}\,,$$

$$e^{V_0^*} = \frac{e^{\alpha p_{nj} + \gamma_j + \varepsilon_{nj}}}{e^{-\alpha\left(c + fee_{nj}\right)/D} - 1}\,,$$

$$V_0^*\left(p_{nj},\varepsilon_{nj}\right)=\alpha p_{nj}+\gamma_j+\varepsilon_{nj}-\ln\left(e^{-\alpha\left(c+fee_{nj}\right)/D}-1\right). \tag{6}$$

The threshold where at least one of the sites is chosen over nonparticipation is given by $V_0^{**}\left(p_n,\varepsilon_n\right)=\max_j\left\{V_0^*\left(p_{nj},\varepsilon_{nj}\right)\right\}$. The threshold V_0^{**} is useful for estimation because simulation techniques will be required only for the portion of $f(V_0)$ consistent with positive demand for a license. Nonparticipation can be estimated using the cumulative distribution of $f(V_0)$, denoted by $F(V_0)$. The revised form for expression (4), giving the probability that individual n chooses alternative k, becomes

$$P_{n0}=1-F[V_0^{**}\left(p_n,\varepsilon_n\right)], \tag{7}$$

$$P_{n,k>0}=\int_{\varepsilon_1=-\infty}^{\infty}\cdots\int_{\varepsilon_J=-\infty}^{\infty}\int_{V_0=-\infty}^{V_0^{**}(p_n,\varepsilon_n)}I[\max_j\left\{U_{n,j>0}\right\}=U_{nk}]f\left(V_0|z_n\right)dV_0 f\left(\varepsilon\right)d\varepsilon_1\ldots d\varepsilon_J . \tag{8}$$

Demographic variables z_n are interacted with the mean of $f(V_0)$ such that $\overline{V}_{n0}=\overline{V}_0+\beta_z z_n$. Note that the upper limit of integration V_0^{**} insures that $\max_j\left\{U_{n,j>0}\right\}=\max_j\left\{U_{nj}\right\}$. The likelihood

---

$$V_0^*=\alpha p_{nj}+\gamma_j+\varepsilon_{nj}-\ln\left(e^{-\alpha\left(c+fee_{nj}\right)/D}-1\right).$$

The final expression is defined as long as the price coefficient α is negative and the choke constant c is positive, both of which are required for consistency with the theoretical model.

function is estimated over all sites and all individuals, including both participants and nonparticipants. The form for the likelihood function is $\prod_{nj}\left(P_{nj}\right)^{y_{nj}}$, where observation $y_{nj} = 1$ if individual n purchases a license at site j and zero otherwise.

There is no closed-form solution for the probabilities in (8) so simulation methods are required to maximize the likelihood function. The normal distributions for ε and V_0 do not carry over into U_{nj}, so probit simulation methods cannot be applied. Instead we use a logit-smoothed accept-reject simulator, an approach that can be applied to any discrete-choice model. It operates as follows: 1) take R draws from the densities $f(\varepsilon)$ and from $f\left(V_0 | z_n\right)$ below $V_0^{**}\left(p_n, \varepsilon_n\right)$; 2) for draw r set $I_{nk}^r = 1$ if $\max\left\{U_{n,j>0}\right\} = U_{nk}$; and 3) calculate $P_{nk}\left|\left(V_{n0} < V_0^{**}\left(p_n, \varepsilon_n\right)\right) = \sum_r I_{nk}^r / R\right.$. Logit-smoothing is applied in step (2) to insure selection probabilities are never zero for any alternative (see Train 2003). Finally, the conditioning on V_{n0} is removed by calculating P_{nk} = $\left\{P_{nk}\left|\left(V_{n0} < V_0^{**}\left(p_n, \varepsilon_n\right)\right)\right.\right\}$ $F[V_0^{**}\left(p_n, \varepsilon_n\right)]$. For every site given each town of origin, P_{nk} was estimated using 2,000 random draws from $f(\varepsilon)$ and from $f(V_0)$ below V_0^{**}.

Results

This section presents results of a "basic" model in which error terms ε_{nj} associated with site constants γ_j are independently and identically distributed. The basic model is analogous to a standard participation model with nesting of sites relative to alternative activities but without nesting among sites (*e.g.*, Parsons, Jakus, and Tomasi 1999). Complete model results are reported only for the basic model, while model extensions that include more general distributions for $f(\varepsilon)$ will be explored in the next section.

Estimated parameters for the basic model are reported in Table 2. The travel-cost parameter is negative and significant. Site constants reflect the greater and lesser desirability of the various sites. The most desirable sites for shellfishing are Duxbury, Wareham, Marion, Mattapoisett, Westport and Rhode Island. The site constant for New Bedford is set to zero for model identification. The standard deviation of site utilities σ_ε is smaller than the standard deviation of V_0, suggesting that substitution is greater among shellfishing sites than between shellfishing and alternative activities.

[Table 2]

The mean utility of alternative activities $\overline{V}_0$ is considerably higher than the site constants, indicating that most individuals take considerably fewer than D trips in a year. Note that for estimation D was set to 60, providing a reasonable upper bound on annual trip demand based on conversations with local resource managers. The estimated choke constant c determines the choke price on individual demands absent a license fee. Equations (2) and (3) show that expected demand associated with c is 0.82 trips per year, which is reasonably close to one, as anticipated. The coefficient on the interaction of $\overline{V}_0$ with income is positive, indicating that all else constant those with higher income are less likely to purchase a shellfishing license. Those with a college education and those with children under the age of 18 are more likely to engage in shellfishing, all else constant.[32]

Model predictions appear in Table 3. Predictions of total license demand for each site correspond reasonably well to data on actual purchases. A comparison to Table 1 shows that nonresident license purchases are somewhat under-predicted for Duxbury and over-predicted for most other sites. For all sites combined, nonresident licenses are overestimated by 18 percent, resident licenses are underestimated by 6 percent and senior licenses are underestimated by less than one percent. Total license demand is predicted almost exactly.

[Table 3]

To further evaluate model results, figures are reported for expected trips per license and value per trip. Expected demand is estimated using equation (2) integrated over the preference distribution for predicted license holders. Predictions range from 1.2 trips per license in New Bedford to 6.1 trips per license in Rhode Island. For all sites combined, average expected demand is 2.0 trips per license. These estimates are reasonably similar to actual demand according to local officials at some sites, while at other sites actual demand may be higher.

[32] A fourth demographic variable identifying residents over the age of 65 was not found to be significant. The distinction could be especially important in coastal towns where demographic factors might confound estimation of the response in demand to lower license fees for resident seniors. However, in specifications that examined the issue, a dummy variable for seniors had a p-value of 0.30 or greater. The variable was therefore dropped from the model. Even without a response to demographic factors, demand is estimated separately for the senior population in each of the coastal towns to explicitly account for the lower senior fees. Data on the size of the senior population by town was obtained from www.umass.edu/miser.

Table 3 includes two estimates of per-trip value in 2004 dollars. Both are calculated using equations (2) and (3). In each case the total change in value across all sites due to loss of access at an affected site is divided by expected demand for the affected site under baseline conditions. This calculation accounts for substitution to other sites in response to the site closure. "Consumer surplus per trip" is calculated net of license fees as specified in (3). Values range from \$5.80 in Rhode Island to \$11.10 in Wareham, with an average of \$10.00 per trip for all sites combined. While consumer surplus per trip is the standard figure reported in the literature, one could argue that an appropriate measure of recreation benefits should not exclude value associated with license fees simply because it has been transferred from the consumer to the licensing authority. This is especially relevant in the case of Massachusetts shellfishing given that annual license fees are as high as \$120 at several popular sites. Figures for "total surplus per trip" include both consumer surplus and license revenue, resulting in a higher per-trip value of \$21.40 for all sites combined.

One complication of this alternative welfare calculation is evident in the wide range of estimated per-trip values, including a negative value reported for New Bedford. The closure of New Bedford leads to a reduction in deadweight loss as recreators switch to more desirable sites previously avoided due to high fees. While unusual in the literature, this result arises from economic circumstances and it is appropriate for welfare estimation to address it. A comparison to day-trip values for other types of water-based outdoor recreation suggests that the \$21.40 figure may be reasonable. For example, a benefit transfer analysis by Rosenberger and Loomis (2001) reports an average value of \$16.37 for swimming and \$31.16 for recreational fishing in the northeastern United States, including Massachusetts.

Model Extensions

The basic model described above assumes error terms ε_{nj} are independently and identically distributed, an assumption that is criticized in the context of discrete choice analysis because it imposes restrictive substitution patterns (Layton 2000; Train 1998). This section explores more general structures for the covariance matrix.[33] The resulting models are described as a "nested" model, a "generalized nested" model and a "random parameters" model because these terms identify analogous specifications in the context of widely used logit methods.

Table 4 reports selected parameters for the three model extensions, with site constants omitted to save space. The "full model" parameters, which apply to all sites in each model, are similar across specifications. To avoid proliferation of parameters, refinements to the covariance matrix are limited to sites along Buzzards Bay, which extends from Wareham to Westport. Similarities in the characteristics of Buzzards Bay sites may lead to correlation in draws of ε_{nj}, corresponding to increased substitution between sites. For example, Buzzards Bay is known to have warmer water than sites above Cape Cod (from Scituate to Plymouth) so a person who prefers warmer water might be represented by a high draw for each of the Buzzards Bay sites. New Bedford and Dartmouth are located on Buzzards Bay but are excluded from the covariance matrix due to difficulties in estimation. The low nonresident license demand at both of these sites appears to be insufficient for estimating substitution patterns. The covariance matrices reported

[33] The full variance-covariance matrix for the model in (1) through (4) can be estimated without imposing restrictions associated with normalization for scale and origin (*e.g.*, probit). Choice utilities are denominated in dollars and therefore fixed in scale. Utilities are normalized for origin because they are implicitly differenced against the utility of nonparticipation, which is constant at zero. In the transformed model, correlation common to all sites must be avoided to prevent redundancy with the variance of V_0.

in the lower portion of Table 4 include (consecutively from upper left) Wareham, Marion, Mattapoisett and Westport.

[Table 4]

The nested model is the least general of the three model extensions. Constraints on $f(\varepsilon)$ are relaxed in the nested model by allowing only identical correlations across sites in the nest. The estimated correlations are positive, consistent with the conjecture that substitution is more likely to occur among sites in the nest. The covariance parameter 0.035 is statistically different from zero, indicating an improvement over the basic model.

[Table 5]

The generalized nested model allows correlations to differ across pairs of sites. This type of model is not common in the literature, but a description of the relevant closed-form logit estimation techniques are given in Train (2003). As in the nested model, the estimated covariance terms are positive and statistically significant. A likelihood-ratio test supports the more general covariance structure.

The random parameters model allows for site-specific variance terms in addition to the full matrix of covariance terms. The variance-covariance patterns presented at the bottom of Table 4 mimic the utility distribution that would be obtained by regressing on site characteristics using random coefficients. The variance terms in the random parameters model for the four Buzzards Bay sites are significantly smaller than variance terms for other sites in the model (for other sites $\sigma_\varepsilon^2 = 0.273^2 = 0.075$). In contrast to the nested models, five out of the six covariance terms are negative. The increase in the likelihood function is again significant.

Table 5 presents a comparison of selected aggregate predictions for the basic model and the three model extensions. The first section of the table reports license demand and welfare predictions for all shellfishing sites combined. The second section shows the effect of model specification on aggregate substitution patterns. Percentages in the second section indicate the proportion of former Mattapoisett license holders who choose each of the available alternatives given loss of access to Mattapoisett. The selection of Mattapoisett for this scenario was based on its high initial demand and central location on Buzzards Bay.

Though many of the figures in Table 5 are not statistically distinguishable from one another, at least two significant trends may be identified. First, substitution patterns shift in the expected direction. The nested model predicts an increase in substitution from Mattapoisett to the other major Buzzards Bay sites (*i.e.*, Wareham, Marion and Westport) compared to the basic model. The pattern is less consistent as the specifications become more general, but substitution within Buzzards Bay is elevated in all three model extensions.

It may seem surprising that the random parameters model predicts elevated substitution within Buzzards Bay given the presence of negative utility correlations. Indeed the negative covariance terms by themselves express differences in the way sites are perceived and valued by a given individual, leading to lower substitution. However, this effect is counteracted by the reduction in variance for Buzzard Bay sites. Reduced variance leads to similar values for utility when compared to high or low realizations for the utility of sites outside Buzzards Bay. For Marion the variance reduction is less pronounced, and the negative correlation with Mattapoisett is significant enough to cause attenuated substitution compared to the basic model. Most models in the literature do not permit negative correlations in site utility (*e.g.*, Herriges and Phaneuf 2002; von Haefen, Massey, and Adamowicz 2005) so it is worth identifying the advantage they provide in this case. Specifically, they allow region-specific participation patterns to be estimated independently of site substitution. For Buzzards Bay sites the reduction in variance corresponds to a more rapid decline in participation across space, because only a small number of utility draws are high enough to offset significant travel costs. The reduced variance simultaneously causes an increase in site substitution, as noted, and negative correlations are needed to appropriately offset this effect.

The second significant trend evident in Table 5 is a shift in the more general models toward a greater number of expected trips per license and a lower value per trip. In the basic model the independent, high-variance distributions for Buzzards Bay preferences generate a large number of high-utility draws. An excess of high draws appears to produce a compensating decline in the choke price on individual demands, forcing the functional form for demand to take a narrower, steeper shape. This would be consistent with lower expected demand per license and higher per-trip value. The prevalence of high-utility draws is mitigated to some degree by correlations in the nested models and by reduced variance in the random parameters model, leading to higher expected demand and lower per-trip value. Support for this explanation is provided by a close correspondence between the identified trends and changes in the choke constant as reported in Table 2 and Table 4.

Conclusions

The preceding analysis has demonstrated the explanatory power of annualized recreation values in predicting discrete-choice behavior. Individual demand functions for expected annual trips, incorporating a choke price to capture nonparticipation and the influence of annual fees, were sufficient to define the utility of choice alternatives in a model of license demand. There are many analogous situations where first-stage choices depend on subsequent recreation behavior. Examples include the purchase of a membership at a tennis club or swimming pool, the selection of a marina by boaters, or the choice of a residence when local recreational opportunities are capitalized into housing prices. Innovations necessitated by the circumstances of Massachusetts shellfishing may prove useful in these other modeling exercises.

References

Bennear, L., R. Stavins, and A. Wagner. 2005. Using Revealed Preferences to Reveal Recreation Benefits: Evidence from Recreational Fishing Licenses. *Journal of Regulatory Economics* 28(2):157-179.

Bockstael, N., I. Strand, K. McConnell, and F. Arsanjani. 1990. Sample Selection Bias in the Estimation of Recreation Demand Functions: An Application to Sportfishing. *Land Economics* 66(1):40-49.

Cesario, F. 1976. Value of Time in Recreation Benefit Studies. *Land Economics* 52(1):32-41.

Englin, J., and J. Shonkwiler. 1995. Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand Under Conditions of Endogenous Stratification and Truncation. *The Review of Economics and Statistics* 77(1):104-112.

Englin, J., P. Boxall, and D. Watson. 1998. Modeling Recreation Demand in a Poisson System of Equations: An Analysis of the Impact of International Exchange Rates. *American Journal of Agricultural Economics* 80(2):255-263.

Forthcoming. Recreation Nonparticipation as Choice Behavior Rather Than Statistical Outcome. *American Journal of Agricultural Economics*.

Gillig, D., T. Ozuna, and W. Griffin. 2000. The Value of the Gulf of Mexico Recreational Red Snapper Fishery. *Marine Resource Economics* 15(2):127-139.

Grogger, J., and R. Carson. 1991. Models for Truncated Counts. *Journal of Applied Econometrics* 6(3):225-238.

Gurmu, S., and P Trivedi. 1996. Excess Zeros in Count Models for Recreational Trips. *Journal of Business and Economic Statistics* 14(4):469-477.

Haab, T., and K. McConnell. 1996. Count Data Models and the Problem of Zeros in Recreation Demand Analysis. *American Journal of Agricultural Economics* 78(1):89-102.

Hausman, J., G. Leonard, and D. McFadden. 1995. A Utility Consistent, Combined Discrete Choice and Count Data Model: Assessing Recreational Use Losses Due To Natural Resource Damage. *Journal of Public Economics* 56(1):1-30.

Hellerstein, D. 1995. Welfare Estimation Using Aggregate and Individual-Observation Models: A Comparison Using Monte Carlo Techniques. *American Journal of Agricultural Economics* 77(3):620-630.

Herriges, J., and D. Phaneuf. 2002. Inducing Patterns of Correlation and Substitution in Repeated Logit Models of Recreation Demand. *American Journal of Agricultural Economics* 84(4):1076-1090.

Huang, J., T. Haab, and J. Whitehead. 1997. Willingness to Pay for Quality Improvements: Should Revealed and Stated Preferences Be Combined? *Journal of Environmental Economics and Management* 34(3):240-255.

Layton, D. 2000. Random Coefficient Models for Stated Preference Surveys. *Journal of Environmental Economics and Management* 40(1):21-36.

Loomis, J., C. Pierce, and M. Manfredo. 2000. Using the Demand for Hunting Licenses to Evaluate Contingent Valuation Estimates of Willingness to Pay. *Applied Economics Letters* 7:435-438.

McConnell, K. 1995. Consumer Surplus from Discrete Choice Models. *Journal of Environmental Economics and Management* 29(3):263-270.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, P. Zarembka, ed., pp. 105-142. New York: Academic Press.

Moeltner, K. 2003. Addressing Aggregation Bias in Zonal Recreation Models. *Journal of Environmental Economics and Management* 45(1):128-144.

Morey, E. 1999. Two RUMs Uncloaked: A Nested Logit Model of Site Choice, and a Nested Logit Model of Participation and Site Choice. *Valuing Recreation and the Environment*, J. Herriges and C. Kling, eds., pp. 65-120. Northampton, MA: Edward Elgar.

Murdock, J. 2006. Handling Unobserved Site Characteristic in Random Utility Models of Recreation Demand. *Journal of Environmental Economics and Management* 51(1):1-25.

Parsons, G. 2004. Travel Cost Models. *A Primer on Nonmarket Valuation*, P. Champ, K. Boyle, and T. Brown, eds., pp. 269-330. Boston: Kluwer Academic Publishers.

Parsons, G., P. Jakus, and T. Tomasi. 1999. A Comparison of Welfare Estimates from Four Models for Linking Seasonal Recreational Trips to Multinomial Logit Models of Site Choice. *Journal of Environmental Economics and Management* 38(2):143-157.

Parsons, G., A. Plantinga, and K. Boyle. 2000. Narrow Choice Sets in a Random Utility Model of Recreation Demand. *Land Economics* 76(1):86-99.

Phaneuf, D. 1999. A Dual Approach to Modeling Corner Solutions in Recreation Demand. *Journal of Environmental Economics and Management* 37(1):85-105.

Rosenberger, R., and J. Loomis. 2001. *Benefit Transfer of Outdoor Recreation Use Values: A Technical Document Supporting the Forest Service Strategic Plan (2000 Revision)*. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Shaw, D. 1988. On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification. *Journal of Econometrics* 37(2):211-223.

Shonkwiler, J., and J. Englin. 2005. Welfare Losses Due To Livestock Grazing On Public Lands: A Count Data Systemwide Treatment. *American Journal of Agricultural Economics* 87(2):302-313.

Smith, K., and W. Desvousges. 1985. The Generalized Travel Cost Model and Water Quality Benefits: A Reconsideration. *Southern Economic Journal* 52(2):371-381.

Sun, L., C. van Kooten, and G. Voss. 2005. Demand for Wildlife Hunting in British Columbia. *Canadian Journal of Agricultural Economics* 53(1):25-46.

Train, K. 1998. Recreation demand models with taste variation over people. *Land Economics* 74(2):230-239.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge, U.K.: Cambridge University Press.

Train, K., D. McFadden, and R. Johnson. 2000. Discussion of Morey and Waldman's 'Measurement Error in Recreation Demand Models.' *Journal of Environmental Economics and Management* 40(1):76-81.

von Haefen, R., M. Massey, and W. Adamowicz. 2005. Serial Nonparticipation in Repeated Discrete Choice Models. *American Journal of Agricultural Economics* 87(4):1061-1076.

von Haefen, R., and D. Phaneuf. 2003. Estimating Preferences for Outdoor Recreation: A Comparison of Continuous and Count Data Demand System Frameworks. *Journal of Environmental Economics and Management* 45(3):612-630.

Table 1

License Purchases and Fees for Shellfishing Sites

	License Purchases				License Fees ($)		
Shellfishing Site	Non-Residents	Residents[1]	Seniors	Total	Non-Residents	Residents[1]	Seniors
Scituate[2]	14	65	59	138	50	20	0
Duxbury[2]	548	422	137	1,107	100	20	0
Kingston	67	109	54	230	55	25	10
Plymouth	18	539	155	712	50	10	0
Wareham	172	984	359	1,515	120	30	15
Marion	100	283	65	448	120	25	0
Mattapoisett[2]	186	350	232	768	120	25	0
New Bedford	3	143	101	247	50	12	3
Dartmouth[2]	13	207	123	343	75	15	0
Westport	65	477	129	671	100	25	10
Rhode Island[3]	46	-	-	46	200	-	-
Total	1,232	3,579	1,414	6,225	-	-	-

[1] "Residents" refers to local inhabitants below the age of 65. "Senior" residents are 65 and older.

[2] In several towns the division of local license purchases into senior residents and residents under 65 is an estimate because precise records were not maintained.

[3] License purchases are not reported for residents of Rhode Island because the target population for the study consists of Massachusetts residents only.

Table 2

Model Parameters

Variable	Estimate	St Err
Travel cost (α)	-0.021	(0.001)
Scituate (γ_j)	0.267	(0.017)
Duxbury	1.019	(0.008)
Kingston	0.480	(0.011)
Plymouth	0.375	(0.006)
Wareham	1.127	(0.012)
Marion	0.920	(0.013)
Mattapoisett	1.032	(0.012)
Dartmouth	0.469	(0.013)
Westport	0.913	(0.006)
Rhode Island	1.398	(0.018)
St Dev of site utilities (σ_ε)	0.195	(0.005)
Alternative activities ($\overline{V}_0$)	5.980	(0.058)
St Dev of V_0	0.647	(0.012)
Choke constant (c)	41.00	(10.30)
Income x $\overline{V}_0$	0.065	(0.004)
Education x $\overline{V}_0$	-0.727	(0.080)
Kids x $\overline{V}_0$	-0.699	(0.099)
Log likelihood	-33,109	

Table 3

Model Results

Shellfishing Site	License Purchases: Non-Residents	Residents	Seniors	Total	Expected Trips Per License	Consumer Surplus Per Trip ($)	Total Surplus Per Trip ($)
Scituate	29	47	42	118	1.5	7.7	11.4
	(4.1)	(3.6)	(2.4)	(9.6)	(0.06)	(0.19)	(0.51)
Duxbury	444	644	210	1,298	2.2	10.1	25.7
	(19.3)	(18.6)	(5.8)	(33.6)	(0.08)	(0.24)	(0.71)
Kingston	104	62	25	191	2.0	6.3	13.7
	(8.7)	(3.1)	(1.0)	(12.4)	(0.07)	(0.16)	(0.55)
Plymouth	33	498	138	668	1.3	9.1	9.2
	(2.6)	(14.4)	(4.0)	(20.0)	(0.05)	(0.23)	(0.38)
Wareham	218	912	331	1,461	2.2	11.1	25.6
	(12.8)	(21.0)	(7.7)	(34.8)	(0.07)	(0.25)	(0.75)
Marion	137	197	123	457	1.8	10.3	18.3
	(9.3)	(7.1)	(3.7)	(17.1)	(0.07)	(0.28)	(0.59)
Mattapoisett	289	288	184	760	2.1	10.4	21.8
	(14.9)	(9.0)	(4.9)	(25.5)	(0.08)	(0.28)	(0.62)
New Bedford	0	173	101	274	1.2	6.8	-0.4
	(0.0)	(9.6)	(5.2)	(14.7)	(0.04)	(0.20)	(0.57)
Dartmouth	13	213	122	348	1.4	9.8	12.3
	(2.2)	(10.9)	(5.0)	(17.2)	(0.05)	(0.27)	(0.44)
Westport	131	340	129	601	2.1	9.9	22.1
	(8.8)	(10.8)	(3.8)	(21.8)	(0.07)	(0.25)	(0.73)
Rhode Island	56	-	-	56	6.1	5.8	27.7
	(6.2)	-	-	(6.2)	(0.19)	(0.15)	(0.83)
Total/Average	1,453	3,374	1,405	6,232	2.0	10.0	21.4
	(34.3)	(48.8)	(25.3)	(82.4)	(0.07)	(0.24)	(0.64)

Note: Standard errors appear in parentheses based on a parametric bootstrap procedure using 200 draws.

Table 4

Parameters for Model Extensions

I. Full-model parameters, omitting site constants

Variable	Nested Model Coeff	Nested Model St Err	Generalized Nested Model Coeff	Generalized Nested Model St Err	Random Parameters Model Coeff	Random Parameters Model St Err
Travel cost (α)	-0.027	(0.002)	-0.024	(0.001)	-0.027	(0.001)
St Dev of site utilities (σ_ε)[1]	0.275	(0.013)	0.267	(0.013)	0.273	(0.009)
Alternative activities ($\overline{V}_0$)	6.259	(0.172)	6.133	(0.083)	6.493	(0.082)
St Dev of V_0	0.743	(0.045)	0.685	(0.018)	0.794	(0.019)
Choke constant (c)	31.47	(11.31)	35.44	(10.44)	30.03	(10.44)
Income x $\overline{V}_0$	0.067	(0.007)	0.066	(0.006)	0.070	(0.005)
Education x $\overline{V}_0$	-0.786	(0.108)	-0.739	(0.102)	-0.935	(0.105)
Kids x $\overline{V}_0$	-0.668	(0.135)	-0.688	(0.116)	-0.889	(0.134)
Log likelihood	-33,085		-33,067		-33,023	

II. Buzzards Bay covariance parameters (Wareham, Marion, Mattapoisett, Westport)

Nested model

$$\Omega_\varepsilon = \begin{pmatrix} 0.075\ (0.006) & \cdot & \cdot & \cdot \\ 0.035\ (0.007) & 0.075\ (0.006) & \cdot & \cdot \\ 0.035\ (0.007) & 0.035\ (0.007) & 0.075\ (0.006) & \cdot \\ 0.035\ (0.007) & 0.035\ (0.007) & 0.035\ (0.007) & 0.075\ (0.006) \end{pmatrix}$$

Generalized nested model

$$\Omega_\varepsilon = \begin{pmatrix} 0.071\ (0.005) & \cdot & \cdot & \cdot \\ 0.023\ (0.006) & 0.071\ (0.005) & \cdot & \cdot \\ 0.040\ (0.005) & 0.030\ (0.005) & 0.071\ (0.005) & \cdot \\ 0.050\ (0.006) & 0.058\ (0.007) & 0.045\ (0.006) & 0.071\ (0.005) \end{pmatrix}$$

Random parameters model

$$\Omega_{\varepsilon} = \begin{pmatrix} 0.018\ \ (0.005) & \bullet & \bullet & \bullet \\ -0.008\ \ (0.002) & 0.046\ \ (0.007) & \bullet & \bullet \\ -0.016\ \ (0.001) & -0.016\ \ (0.005) & 0.031\ \ (0.004) & \bullet \\ -0.011\ \ (0.002) & 0.028\ \ (0.005) & -0.002\ \ (0.004) & 0.026\ \ (0.004) \end{pmatrix}$$

---

[1] The standard deviation σ_{ε} applies to all sites except for Wareham, Marion, Mattapoisett and Westport, which are described by the covariance matrices in Section II of the table.

Table 5

Comparison of Model Results

Statistic	Basic Model Estimate	Basic Model St Err	Nested Model Estimate	Nested Model St Err	Generalized Nested Model Estimate	Generalized Nested Model St Err	Random Parameters Model Estimate	Random Parameters Model St Err
I. Demand and welfare predictions (all sites combined)								
Nonresident licenses	1,453	(34.3)	1,500	(37.3)	1,493	(43.0)	1,497	(36.2)
Resident licenses	3,374	(48.8)	3,295	(72.1)	3,322	(43.5)	3,297	(48.1)
Senior licenses	1,405	(25.3)	1,432	(40.6)	1,415	(24.2)	1,444	(29.2)
Total licenses	6,232	(82.4)	6,228	(87.9)	6,231	(78.5)	6,238	(80.0)
Expected trips per license	2.0	(0.07)	2.3	(0.10)	2.2	(0.11)	2.3	(0.08)
Consumer surplus per trip ($)	10.0	(0.24)	9.0	(0.27)	9.4	(0.30)	8.9	(0.22)
Total surplus per trip ($)	21.4	(0.64)	19.1	(0.73)	20.2	(0.81)	18.3	(0.54)
II. Substitution patterns: Given loss of access to Mattapoisett, former Mattapoisett license holders choose								
Wareham (%)	7.5	(0.40)	11.0	(0.96)	10.4	(0.88)	8.8	(0.72)
Marion (%)	25.6	(0.97)	28.4	(2.29)	25.4	(0.90)	24.1	(1.20)
New Bedford (%)	0.0	(0.00)	0.0	(0.01)	0.0	(0.02)	0.1	(0.01)
Dartmouth (%)	1.2	(0.16)	1.0	(0.21)	1.0	(0.24)	1.3	(0.18)
Westport (%)	5.6	(0.30)	6.9	(0.56)	8.1	(0.92)	8.6	(0.76)
Non-Buzzards Bay sites (%)	6.3	(0.24)	4.2	(0.30)	3.6	(0.33)	6.7	(0.37)
Other activities (%)	53.9	(1.20)	48.4	(3.36)	51.4	(1.21)	50.5	(1.44)
Total (%)	100.0	-	100.0	-	100.0	-	100.0	-

Note: Standard errors in parentheses are based on a parametric bootstrap procedure using 200 draws.

Value of Information from Soil Surveys

Archana Pradhan and Jerald J. Fletcher
PhD Candidate and Professor of Resource and Environmental Economics
Division of Resource Management
West Virginia University

June 30, 2008

This research was supported by the West Virginia Agricultural and Forestry Experiment Station, Natural Resources Analysis Center, and the USDA-NRCS, National Geospatial Development Center under Cooperative Agreement # 68-3-D47-24-2

The Value of Information from Soil Surveys

Abstract

The National Cooperative Soil Survey (NCSS) is the primary source of information on soils in the United States. The benefits attributable to the NCSS information are diverse and dispersed spatially, temporarily, and among user groups. Implementation of the NCSS as a cooperative program between federal and state governments provides a natural experiment to determine the impacts of the NCSS information. These data are analyzed to find evidence for and estimates of past benefits through an analysis of corn production using county-level data in the Corn Belt. The statistical results are strong; initial analyses indicate corn yield increases of 2.5 bushels per acre per year are attributable to the provision of soils information.

Keywords: Soil Survey, Benefit estimation, corn yield

Introduction and Research Objectives

Problem Statement

The National Cooperative Soil Survey (NCSS), a cooperative effort of federal, state, and county agencies, is the primary source for collecting and providing soil data for the United States. The NCSS carries out its activities on national, regional and state levels under the leadership and coordination of the U.S. Department of Agriculture (USDA), Natural Resources Conservation Service (NRCS). The soil survey program was formally initiated in 1899 with the first report of field operations, USDA Report 64, published by the USDA Division of Soils (Smith, 1998).

There is little knowledge of the value of or the benefits derived from the information provided by soil surveys. Since the comprehensive soil survey program is nearing completion, it is appropriate to review the contributions of the program. Estimates of the values of historical and current benefits as well as potential future benefits from additional investments would facilitate program management and guide policy decisions that will determine future investments.

This paper analyzes the effects of the availability of soil survey information on corn yield trends for primary corn producing counties of the U.S. The estimates obtained from this study will serve as a partial measure of the benefits due to the increased availability of soils information provided by the NCSS program. Value of the increased corn yield attributable to the availability of soil information provides an estimate of the economic benefits generated by the use of the soil information in corn production. These benefits depend on changes in management activities, land acquisition patterns and choices on the extensive margin that result from increased availability of soil information. To our knowledge, this is the first attempt to elicit statistically reliable and valid estimates of the benefits of soil survey information on aggregate agricultural production.

Background

Users of soil survey information include farmers, ranchers, foresters, planning agencies, researchers, engineers, development organizations and private investors. Farmers use soil information to manage, expand and to select appropriate farming techniques. Foresters use soil information for selection of sites for plantations, selecting tree species which vary in productivity by soil characteristics and for management activities. Unlike farmers and foresters, planning agencies are focused on broader uses, such as agriculture to urban land and grazing land to forest land conversions. Engineers use soil information to evaluate construction sites, plan road alignments, design building foundations and evaluate sewage disposal potential.

Soil survey information is used in such diverse fields as farming, ranching, forestry, urban planning and zoning, site selection for buildings, roads, airports, recreational sites and parks, and for other purposes as well. Thus the benefits from the soil survey are diverse and dispersed spatially, temporarily and among user groups. Some of the benefits are immediate, some are realized over time and some are only realized over a longer period. Aggregating economic values for a program that provides such varied and diverse benefits is complex.

The primary goal of the soil survey program is to assist society and individuals to understand the suitability and limitations of the soil resources for intended uses (Ditzler, Engel and Ahrens, 2003). Soil surveys provide information that allows users to predict the consequences of alternative uses. Young (1973) stated that the “main purpose of soil surveys is to serve as a basis for decisions concerning land-use”. Likewise Bie et al. (1973) pointed out that producers receive optimal returns when land use and management are adapted appropriately to local soil conditions. Soil information has been used for centuries to guide farmers to manage and better understand crop growth (Samuelson et al., 2002).

Soil maps and the attributes of the various soil series derived from the soil survey provide information to farmers for site selection, land use and management activities. Farmers use soil information to determine the capability of soils to sustain certain kinds of crops, the relative productivity of farm fields, and the best production practices for a given situation. Soils information can thus affect agricultural production related decisions on both the intensive and extensive margins. On the intensive margin, soil information affects crop and rotation choice as well as fertility, tillage, and other management activities. On the extensive margin, soil information affects land purchase decisions and the movement of marginal lands in and out of production.

Klingebiel (1966) reported that farmers in Hall County, Nebraska had extra income because of the availability of soil information which helped them to improve water management and reclaim saline land. He also found in one case that the income of a farmer in Fayette County, TN was increased by more than $5500 in single year as a result of management changes related to soil information. Thus, production agriculture is one sector that benefits directly from the availability of soil survey information.

Studies by Garcia et al. (1987) and Menz and Pardey (1983) found that the increasing trend in corn yields in the U.S. was primarily due to the adoption of new agricultural technologies. Schroder et al. (1984) noted that any issues related to the contribution of specific technologies to changes in agriculture production assume something about the underlying production function. Estimation of benefits of technology or other factors that contribute to agriculture production can be analyzed within the explicit production function framework at the farm or field level or through supply and demand analysis at aggregate levels. This study relies on an aggregate analysis based on county level corn yield trends to quantify the benefits attributable to improved soil survey information.

Research Objectives

The objective of this paper is to develop methods to estimate the increase in corn production due to availability of soil information in major corn producing counties, which includes identifying the data and appropriate statistical techniques and using the estimates to develop measures of value.

Valuation of Soils Information

Valuation of Soils Information

This section provides a review of the theoretical issues behind the concept of the "economic value of information" to provide a background and develop a theoretical foundation for the analysis that follows.

Value of information

According to information theory, information is defined as reduction of uncertainty. According to McGee and Prusak (1993), information can be considered as data, both factual and numerical, that is organized and imbued with meaning as a result of gathering, analyzing or summarizing the data in a meaningful way.

The value of information is an outcome of choice in uncertain conditions (McCall, 1982). It is the difference between the project value with the information and the project value without the information, minus the cost of acquiring the information. The value of information is determined by its importance to the decision makers or to the outcome of the decision. Decision makers may be willing to pay for information depending on the degree of uncertainty and what is at stake (Macauley, 2005). Information has value when the alternative outcome can be different; otherwise information has no role in adding value. In other words there must be uncertainty, and if there is uncertainty, there must be choices. If there are no choices, there are no decisions to be made, and information has no value.

More information helps individuals make better decisions. Better decisions increase expected utility or decrease expected cost. Individuals can be expected to be willing to pay for additional and improved information if the cost of the information is lower than the expected value of their gains.

Macauley (2005) specifies that the value of information depends on the following factors:

1. Degree of uncertainty of the decision maker: How much information will help in making the decision? If there are few actions available, information can have low value.
2. What is at stake (value of output as an outcome of the decision): Value of output is the total value of resources or activities as an outcome of the decision. Willingness to pay for information is a derived demand. How much could the final value of the output be affected?
3. Cost of information to make the decision.
4. Price of substitutes for the information: Are there any alternatives? If so, what is the cost for the substitute?

The larger the degree of uncertainty and the value of output, the larger the value of information; the larger the cost of information and the lower the price of substitutes, the smaller the value of the information will be.

Soil information

The soil survey program, formally initiated in 1899, initially concentrated on the capabilities of land for agriculture production. Soil information has traditionally been presented as maps showing the distribution of soils in a particular area. Properties used in classifying soils include, but are not limited to: soil texture (grain size, color), organic matter content, moisture content, permeability, slope, elevation and water holding capacity. Some soil information based on a one time sample is valid for many years, e.g. elevation, landscape position, texture and density. In contrast, regularly sampled data that reflects temporally varying information includes characteristics such as moisture content, ground water level, soil acidity, nitrogen content, phosphorous content and the potassium level. Soils with similar properties are grouped in mapping units. The U.S. programs are normally conducted at the county level; after completion of the survey in a county, the soil survey is published to provide soil information to the public. Experts in soil science can use soils information to predict the response of specific soils to various uses and management activities.

Soil surveys are classified into five orders from the first to the fifth based on the intensity of field study, the degree of mapping detail, the phase or levels of abstraction in defining and naming map units and different map unit designs (Soil Survey Division Staff, 1993). Figure 1 represents the soil geography hierarchy as a reverse pyramid proceeding from the most general at the top to the most specific at the bottom. The NCSS program supports second-order surveys that are nearly complete for all private lands in the U.S. and represent cooperative efforts between state and county governments and the USDA/Natural Resource Conservation Service (see the Soil Survey status map in Figure 2).

The information provided by the NCSS has played a significant and important role in diverse fields. There are considerable challenges to fully estimate the aggregate benefits derived, but such estimates are needed to conduct an accurate cost-benefit analysis.

Temporally, benefits provided by soil surveys can be broadly categorized into the following three groups:

1. Historical benefits (benefits achieved in the past).
2. Current benefits (benefits realized in the current period from the use of NCSS developed soil survey information).
3. Future benefits (benefits expected in future years from the availability of NCSS soil survey products as well as continuing activities).

Historical and current benefits are derived from past investments. Future benefits can be further divided in two categories: 1) benefits to be derived from past investments in the NCSS program, and 2) benefits that will be derived from additional (current and future) investments in the NCSS soil survey program.

Estimates of past benefits provide a measure of the returns to past investments. Current benefits give a measure of the ongoing returns to past investments. All types of benefit estimates depend on time, duration, uses and the user groups considered. Some of the benefits realized in the past and continuing to the current period can be estimated through indirect methods using currently available data. Partial future benefits can be estimated by extrapolating from such analyses.

Economic Analysis of Availability of Soil Information

Policies having nonprice effects on the producer must sometimes be evaluated (Just, Hueth and Schmitz, 2004). The government has made investments in collecting and providing soil information for the public good. This information, provided to users free of charge, substantially affects aggregate productivity (Just, Hueth and Schmitz, 2004). To account for nonprice impacts on producers, the interpretation of fixed factors of production can be expanded. Since such factors do not exist in markets, demand for such factors is not directly observable.

Benefits derived from soil survey information can be demonstrated using a standard supply and demand framework and economic welfare methods. Supply of any good depends on price and production cost as well as other factors. For instance, introduction of soil information may change the supply curve of a particular crop. An increase in supply can result from adjustments on both the intensive and extensive margins. Introduction of soil information helps farmer to better understand and manage their land which can increase yield. Soil information also affects land purchase decisions and stimulates movement of marginal lands in and out of production, which also affects crop supply curves.

The benefits derived from the information provided by soil surveys and the cost to produce the soil surveys can be computed using generally accepted welfare economics methods. Welfare economics is based on the idea that a change in an individual's economic well-being can be measured in terms of the individual's willingness to pay to obtain the change (in case of a good) or willingness to pay to avoid (in case of a bad). All individuals in society are categorized as producers, consumers or both in order to analyze changes in social welfare in market terms. Consumers' welfare is measured by using consumer surplus (as a first approximation) while producers' welfare is measured using producer surplus. In Figure 3, suppose D is the demand curve and S_0 the initial supply curve for a crop (e.g. corn). The area below the demand curve and above the initial price, P_0, given by the initial supply curve, S_0, represents consumer surplus (area ABP_0). Producer surplus is the area above the supply curve and below the price line, P_0, (area P_0BC).

Assuming the change in supply of the crop due to provision of soil information reflects true social value, the welfare effects are represented by Figure 3. The initial equilibrium, i.e., before the availability of soil information, is represented by point B, the point that generates maximum social welfare, i.e., the sum of producer and consumer surplus, when farmers did not have soil information. The use of soil survey information helps farmers better manage their land and affects decisions regarding the inclusion of land in crop production. Thus introduction of soil survey information increases yield and reduces the marginal and average costs of production, shifting the supply curve outwards from S_0 to S_1; this results in a new equilibrium at point D. This results in an overall gain in social welfare equal to area $BDEC$. Because of the reduction in price due to higher output, consumers unambiguously gain an amount equal to area P_0BDP_1. Producers gain an area of P_1DE less P_0BFP. The producers' net gain from the introduction of soil survey information is ambiguous, depending on the relative elasticity of supply and demand. If the demand is elastic, producers are likely to gain. However, if the demand is more inelastic, producers are most likely to lose.

Nonmarket Valuation Approaches to Valuing Soil Survey Information

Soil survey information in the United States is considered public property (Soil Survey Division Staff, 1993). Goods and services that are both nonexcludable and nonrival are public goods. Nonexcludable means that no one can be excluded when the good is provided. Nonrival means that one person's consumption does not reduce the ability of others to consume that good. Since soil information is potentially excludable, it can be defined as quasi-public goods although current policies make the information closer to a true public good.

Because of the public good nature of soil survey information, the economic value of the soil survey is not directly observed in market transactions. It is thus difficult to estimate the economic value from additional investment. In these cases, non-market valuation techniques could be applied to estimate the economic value of soil information that society receive from uses of soil survey information. The value of public goods can be measured as willingness to pay (WTP) and willingness to accept (WTA) using direct value elicitation methods. WTP is the maximum amount of money an individual is willing to pay for the improvement (additional investment) and WTA is the minimum amount of money the individual would require to forgo the improvement (Freeman, 2003). In the case of additional investments in the soil survey, WTP is the compensating variation measure of welfare change, whereas WTA is the equivalent variation.

Two approaches could be employed to estimate the benefits provided by soil information:

1. Direct methods
2. Indirect methods

Direct methods are survey based approaches to valuation usually based on individual responses. Such methods attempt to determine the value for a public good by directly asking individuals. The contingent valuation approach is a commonly used direct method based on the decision maker's responses to hypothetical questions. Properly constructed, such surveys provide the information needed to conduct traditional demand analysis. It is one of the oldest methods to elicit consumers WTP for nonmarket goods (Young, 2005). Mitchell and Carson (1989) argue that the contingent valuation method is the most promising approach for determining WTP for many public goods, if the method is applied carefully.

Indirect methods involve observing real world behavior in response to a public good and then applying economic models and statistical analysis that allow us to extract and identify the value of the public good. Indirect methods rely on statistical procedures within an accepted economic framework to capture the impacts on decisions and related outcomes. The analysis of econometrically estimated production and demand functions provides an example of the use of indirect methods. The production approach begins by trying to measure the contribution of the public good to output derived from its use through standard aggregate production relationships that depend on a vector of standard factors of production in addition to the soil information. For instance, the impact of soil survey information on aggregate corn production using a panel data approach of average county corn yield over time for several hundred counties can capture the impacts of temporally distributed access to soil survey information in aggregate production functions.

The general form of the production function can be expressed as output is a function of a vector of factors that contributes to output in addition to the soil survey information:

$$Y = f\ (S, X)$$

Where Y represents average corn yield, S represents soil information and X represents the set of variables that determine corn yield trend such as technology, hybrid, weather, fertilizer and pesticides. The effect of the introduction of soil survey information or a change in soil survey information for a county can estimated by measuring the impact on corn yield correlated with the provision of information that is not explained by the usual inputs. The final form can be manipulated to isolate the impacts of soil information.

This study relies on indirect methods to analyze econometrically estimated production relationships to measure the value of soil information. The primary development is through a case study of corn production in the Corn Belt. The development of soil survey information over the past 60 years in conjunction with the data available on corn production provide the results of a natural experiment that can be analyzed to evaluate the impacts of the NCSS on corn production.

Literature Review

Previous Studies of the Benefits of Soil Survey Information

There has been little research conducted on the benefits derived from the NCSS program. Western (1978) defined the soil survey value as the ratio of survey quality to survey cost. Bie et al. (1973) and Beckett and Burrough (1971) noted that the cost and benefits of soil survey rises sharply with increasing quality.

Klingebiel (1966) estimated cost benefit ratios for soil survey investments according to the intensity of land use: 1) low intensity (predominantly range and woodland), 2) medium density (mixed agriculture, and about half cropland), and 3) high intensity (rapidly growing metropolitan areas). He showed that the benefits of soils information increase with increasing land use intensity. He estimated a cost benefit ratios for low intensity areas of 1:46, for medium intensity areas of 1:61 and for high intensity area of 1:123. Estimates of the benefits were determined on the basis of the case histories and record of soil survey users, assuming that most people in the surveyed area would use soil information.

Bie and Ulph (1972) showed that the value of soil survey information depends on the quality of the maps developed and differences in payoffs among alternative management practices. Their study, based on varieties of peaches, showed that gross returns increase as the quality of the information of each mapping unit increases. Dent and Young (1981) noted that U.S. and Australian studies demonstrated benefit-cost ratios from 40:1 to 50:1. They used a straightforward methodological approach to estimate the economic benefit of a soil survey by comparing the profitability from different management systems on each of a number of mapping units.

More recently, Giasson et al. (2000) derived an estimated economic value of \$17.14/hectare each year from a hypothetical soil survey, which exceeded the estimated soil survey cost of \$2.09. They used decision trees, Bayes' Theorem, and map quality evaluation procedures to evaluate the economic value of soil surveys from three different scenarios which differed in the level of information concerning soil changes. The three scenarios they considered are: (i) site-specific soil information is unavailable, (ii) perfect site-specific soil information is available (not realistic), and (iii) imperfect site-specific soil information is available.

Previous Corn Yield Studies

A number of researchers have attempted to estimate the influence of weather and technology over a long period using variety of techniques. Generally two types of approaches have been employed to assess the impact of weather on crop yields: crop growth simulation models and statistical models. Most studies have used a model with a single-equation framework (Huff and Neill, 1982; Offutt, Garcia and Pinar, 1987; Kaufmann and Snell, 1997).

Swanson and Nyankori (1979) assessed the impact of weather and technology on yield growth of corn and soybeans on the Allerton Trust Farm, Piatt County, Illinois for the 1950-1976 period by comparing yield trends not adjusted for weather with yield trend adjusted for weather. They used monthly temperature and precipitation data for June, July and August. Their analysis showed that yield increases according to a linear time trend, which serves as a proxy for technology; they found that using various non-linear formulations did not improve the model significantly.

Huff and Neill (1982) expressed yield as a function of time and weather variables in their study of corn yield for regions of the Midwest for 1931-1975. They concluded that July and August temperature and July precipitation are the most important explanatory variables. This corresponds to the relatively short reproductive stage (grain formation period), two to three week period in July in the Midwest and the historical fact that favorable August weather can enhance yield. They found the quadratic trend (including both linear and quadratic time terms) as statistically adequate to represent technological improvements. Thompson (1975) also used a linear and quadratic time trend proxy to represent technological change for 1960 onwards.

Garicia et al. (1987) examined the relationship between yield level and yield stability with respect to advances in technology and weather conditions for corn. They divided the yield data for 1931-1982 into two different sets based on the history of technological advance. Using a linear time trend as a proxy for technological advances, they found that yield behavior adjusted for weather resulted in nearly identical yield variances for two different periods (1931-60 and 1961-82), which suggests that technology is not the only determining factor responsible for yield behavior.

Kaufmann and Snell (1997) estimated a hybrid model accounting for both climate and social determinants of corn yield using data from counties in the eight largest corn producing states for 1969-1987. They found that use of county-level data captured the significant variations in temperature and rainfall occurring within the states (Kaufmann and Snell, 1997). They used a time trend to represent the effect of technological advances and hybrids that could not be measured clearly.

Hu and Buyanovsky (2003) investigated the climate effects on corn yield using data from Sanborn Field in Columbia, Missouri for the 1895-1998 period. Their results indicate that the climate effects can be better explained by within-season variations in temperature and precipitation rather than by average growing season conditions. More recently, Schlenker and Roberts (2006) employed a reduced-form model to relate weather and corn yield using detailed daily weather records for about 800 counties in the eastern United States for 1950-2004. Their results indicate a significant nonlinear relationship between corn yields and temperature. Yield was found to increase with moderate temperatures but the response was not favorable after temperatures exceed 30^{o} C.

Methodology

The model is based on the knowledge that primary production in agriculture is dependent on climate, soil, and the technology in a society. The yield of an agricultural crop, or of any other plant, is governed by the nature of the soil supporting it, weather, and management practices (Simonson, 1955). The model is based on estimating corn yield trends as a function of spatially and temporally varying weather data, time trends that reflect technical and management change, and the time of introduction of soils information by county as soil surveys were completed.

Statistical Model

To our knowledge, none of the previous corn yield studies have included soil information as an explanatory variable. A model integrating soil information with other variables such as technology and weather, can be employed to estimate the contribution of soil information to aggregate corn yield. The general form of this model is expressed as:

Yield = f (soil information, technology, climate)

The wide spread spatially and temporally diverse nature of the provision of soil survey information supports the contention that provision of the soil survey information is not systematically correlated with other variables such as technology, fertilizer use and the introduction of hybrids. If so, the methods developed provide an unbiased estimate of the value of soil information for corn production. The temporal trend is captured in a time technology trend; the primary spatial and temporal variability is captured by the county level weather measures for each year.

The data are developed as a panel data set. The combination of time series with cross-sections enhances the quality and quantity of data in ways that would be impossible using only one of these two dimensions (Gujarati, 2004). Panel data are more informative; provide more variability, less collinearity among variables and more degree of freedom; and give more efficient estimates (Baltagi, 1995). This approach provides control for individual unobserved heterogeneity which is not easily detectable in either cross-section or time-series data. A panel data regression is expressed with double subscripts on variables. The model can be shown as:

$$Y = \alpha + \beta X'_{it} + u_{it} \quad i = 1,.....N; \; t = 1,...T$$

The subscript i denotes the cross-section dimension and t denotes the time-series dimension. In this analysis model i represents counties and t represents years. The error term in panel data analysis are decomposed in two components:

$$u_{it} = \mu_i + \nu_{it}$$

where μ_i denotes the unobservable county specific error and ν_{it} denotes idiosyncratic error. Error term μ_i does not change over time and accounts for any county specific effect that is not included in the regression. Error term ν_{it} varies over counties and year.

Generally two types of models are available for panel data analyses: fixed effects models and random effects models. In the fixed effect model, the μ_i are assumed to be fixed parameters to be estimated and the ν_{it} independent and identically distributed $IID(0, \sigma_\nu^2)$. The fixed effects model consists of too many parameters and suffers from a loss of large degrees of freedom. Loss of degrees of freedom can be avoided if the individual effect μ_i can be assumed to be random. The random effects model assumes the individual effect μ_i is random. In this case both μ_i and ν_{it} are $IID(0, \sigma_\nu^2)$ and μ_i are independent of ν_{it}. Also, the independent variables, X_{it}, are independent of μ_i and ν_{it} for all i and t. The random effects model is appropriate when the individuals are selected randomly from a large population (Wooldridge, 2002). Fixed effect models are usually much more convincing than random effect models for policy analysis based on aggregated data (Wooldridge, 2006). Since all counties from major corn producing states are used in this study, the fixed effect model is employed to estimate the regression equation. The use of a fixed-effects panel estimator allows us to interpret the regression coefficient estimate of increase in corn yield as a measure of soil survey benefits on corn production as soil survey information is available.

The data for this research are stacked into a cross section of time-series before analysis using a fixed-effects panel estimator. Thus the data set comprises *T* observations for each on *N* counties. Formulation of fixed effect model assumes that the variation across counties can be captured in the constant term. Each individual county-specific constant is treated as an unknown parameter to be estimated. The equation being estimated is

$$y_{it} = \alpha_i + \beta X'_{it} + u_{it} \quad i = 1, ...N; \;\; t = 1, ...T$$

where y_{it} is corn yield in county *i* in year *t*, α_i is a county-specific constant (which is allowed to be unique for each county), β is a vector of coefficients, X_{it} is a vector of independent variables, and u_{it} is an error term for each county-year observation. This model is also referred as least squares dummy variable (LSDV) model (Greene, 2003). The least square estimator of β is given by

$$\hat{\beta} = [X^{'} M_D X]^{-1} \; [X^{'} M_D y],$$

where

$$M_D = I - D(D^{'}D)^{-1} D.$$

X is the entire matrix of independent variables including the county-specific intercepts, y is the vector of observations on county yield. This equation sums to a least squares regression using the transformed data $X_* = M_D X$ and $y_* = M_D y$. M_D is symmetric, idempotent and orthogonal to D.

$$\mathrm{M_D} = \begin{pmatrix} M^0 & 0 & 0 & .. & 0 \\ 0 & M^0 & & .. & \\ 0 & & M^0 & .. & \\ & & .. & .. & \\ 0 & 0 & 0 & .. & M^0 \end{pmatrix}$$

In this formula, $M^0 = I_T - ii^{'}/T$, where I_T is an identity matrix of rank T, i is a $T\times1$ vector of ones, and T is the number of periods over which the cross-sections are observed ($T = 72$). Thus, if there are n counties observed for 72 years each and k explanatory variables including the constant and the fixed effects, then X is a $72n\times k$ matrix, y is a $72n\times1$ vector, M^0 is a 72×72 matrix, I_n is a nXn identity matrix, and M_D is a $72n\times72n$ matrix. The matrix M_D controls for correlation across the error terms within counties. The least squares regression of $M_D y$ on $M_D X$ is equivalent to a regression of $\left[y_{it} - \overline{y_{it}}\right]$ on $\left[x_{it} - \overline{x_{it}}\right]$, where $\overline{x_{it}}$ and $\overline{y_{it}}$ are scalar and K x 1 vector of means of y_{it} and x_{it} over T observations for group i.

The county-specific effects α_i will capture all time-invariant characteristics of a location in the above fixed effect model. The use of fixed effects model avoids the problem of omitted variables, since they are included in the fixed effects (Schlenker and Roberts, 2006).

Data Description

USDA-NASS county-level corn yield data from 1935-2006 were obtained for the ten Corn Belt States from the USDA-NASS Web site (http://www.nass.usda.gov/). All the counties for Corn Belt States (Indiana, Illinois, Iowa, Missouri and Ohio) and the neighboring states (Michigan, Nebraska, Wisconsin, Minnesota and South Dakota) were selected (Figure 4). The ratio of corn planted to total crop land was calculated based on 1987 data (Figure 5). The highest ratio is 0.6 and only 13 counties had the ratio greater than 0.5. To avoid the counties with none or very low corn plantation, different sets of data with positive ratios were used in the analysis. The final set of the data were selected with a ratio greater or equal to 0.2.

Soil survey status data is available from the Natural Resources Conservation Service, NRCS (Figure 2). The date when soil survey information was certified and was started to be used was employed as the year for soil information available year. The latest date is 2006 and the earliest date is 1952 for the soil survey certified and began to be used. Dummy variables were created for the soil survey information, with value 0 before the soil survey information was not certified and was started to be used and value 1 after the soil survey was certified and was started to be used.

Based on the previous studies (Kaufmann and Snell, 1997; Garcia et al., 1987; Schroder et al., 1984; Thompson 1975) the possible weather variables that could be used in this study are preseason moisture data, monthly precipitation and temperature for June, July and August. Thus nine weather variables were used in the model which are minimum temperature, maximum temperature and precipitation for June, July and August. Gridded climate data files were obtained from Michigan State Climatologist's Office (Figure 6). The data included longitude, latitude (in hundreths of degrees) and the daily value for the grid point. This was given in each file for all of the grids in the lower 48 states for each year. An inverse distance weighted (IDW) technique was used to interpolate a surface from grid points. A neighborhood about the interpolated point is identified and a weighted average is taken of the observation values within this neighborhood. The weights are a diminishing function of distance. IDW methods are based on the assumption that the interpolating surface should be influenced most by the nearby points and less by the more distant points. Various options are available for IDW interpolation technique. Precipitation records can have a short spatial correlation length scale and large variability, where as the temperature records have a long spatial correlation scale (Shen et al., 2001). Thus, for interpolating the precipitation data more emphasis was given on the nearest points. Temperature and precipitation data was than recorded for each county centriods from the interpolated surface. Monthly weather values were then obtained for each county by averaging the daily values.

Past extreme weather events have caused severe crop damage and consequently caused a significant economic loss. Most of those weather events are captured by the above discussed nine variables, however the effect 1993 Mississippi River Valley floods was not captured by these variables. Flooding in the summer months of 1993 affected 16,000 square miles of farmland in the Midwest damaging over 11 million acres of crops (Rozenzweig, 2001). Thus to capture the effect of this 1993 flood event, a dummy variable for year 1993 was included in the model.

Technology variables based on previous studies (Griliches, 1957; Kaufmann and Snell, 1997; Schroder et al., 1984) that could be employed in this study include hybrid introduction and fertilizer use. However, the information to develop comparable data for these variables across the spatial and temporal dimensions of this study are not available. In the case of hybrids, there are not studies that imply a spatial variation in introduction. After more than three months spent to collect fertilizer information, the lack of consistent information became apparent. The sources to gather and report the fertilizer data vary significantly across both space and time. State level fertilizer information is reported by USDA only after 1966. Alexander and Smith (1990) estimated a county-level nitrogen and phosphorous fertilizer use for the period 1945-1985, by disaggregating state-level USDA data (1966, 1976, 1977-1985) to county-level. However, they noted that county-level estimates of fertilizer use prior to 1970s should be used with caution. Likewise, fertilizer sales data was provided by USGS Water Resources Division for the period of 1986-1991. Since our model employs data from 1935-2006, it is impossible to acquire fertilizer data from the beginning of our study period. Based on our approach, consistent time trends are of utmost importance. Projecting fertilizer use data for the earlier period (before 1966) may impose additional errors and bias our results. While inclusion of fertilizer data is preferred, because of the statistical issues, this research has not included fertilizer data in the analysis.

Thus, time trend variables are included in the model to capture predictable patterns of technological growth. Several past studies have included a time trend as a proxy to estimate the effect of technology on yield (Garcia et al., 1987; Kaylen et al.1992; Houck and Gallengher, 1976: Menz and Pardey, 1983; and Buller, 1972). Linear, square, cube and fourth order polynomial trends are used in the model to disentangle technological effects such as fertilizers, hybrids, and pesticides. Higher-order polynomial trends were also considered, but the improvement in the model was negligible. Thus up to fourth order polynomial form of time trend was selected.

Eviews software was used to perform the calculations. The results of the corn yield model using the fixed effects regression analysis is presented in Table 1. The graph of mean and standard deviations of residuals from the regression equation for soil survey is presented in Figure 7. The nature of residuals does not show any distinct pattern. However, graph of standard deviation shows an expanding pattern as time increases. Even though this might be a problem of heteroscadesticity, absence of bulges or sudden expansion suggests that it might not seriously affect the results.

Results and Limitation

Results

A shift in the long term corn yield trend was expected initially. The regression coefficients for the soil information variable, nine weather variables and time trend are reported in Table 1. The dummy variable of soil survey information was found statistically significant. Likewise, time trend, dummy variable for 1993 flood and all the weather variables except June maximum temperature were found to be statistically significant. All the variables had the expected signs and the signs were same compared to the previous studies (Table 2).

Moran's I index to test spatial autocorrelation of the residuals obtained from using the first soil survey was detected positive (Figure 8). This could be because of the pattern of soil survey published date, which also exhibits a clustered pattern (Figure 9).

Limitations

The estimates calculated in this study are preliminary subject to omitted variables. However, since our main focus in the study is to estimate the contribution of soil information to corn yield, the estimates should provide information to estimate the benefits of soil information to corn production.

To estimate the net benefits of soil survey program, the benefits from the soil survey to other sectors should also be estimated. Aggregating all the benefits temporally and among different user groups is necessary to provide the actual benefits of soil survey information.

Future research focusing on estimating the benefits in other sectors that benefit from soil survey information is desirable. Future benefits could be estimated using other nonmarket valuation approaches. Since soil survey information is a public good, all the benefits are difficult to capture. Other economic tools such as survey based approaches could be useful in capturing some of the future benefits.

Conclusion

The use of natural experiment in this study provided a means of estimating and calculating a more realistic partial evaluation of the value of soil survey information in agriculture production. The result finds that a yield increase of 2.5 bushels per acre each year can be attributed to the provision of soil information. This is a very promising result given the incomplete nature of the currently available data. The result provides information needed to calculate estimates of the economic value of the NCSS soil information for corn production in the Corn Belt. This result combined with estimates of the value of soil information for other uses and in other sectors provides information for policy makers to make decisions on the future of the NCSS program.

Figure 1: Soil geography hierarchy diagram (Source: Soil Survey Division Staff, 1993)

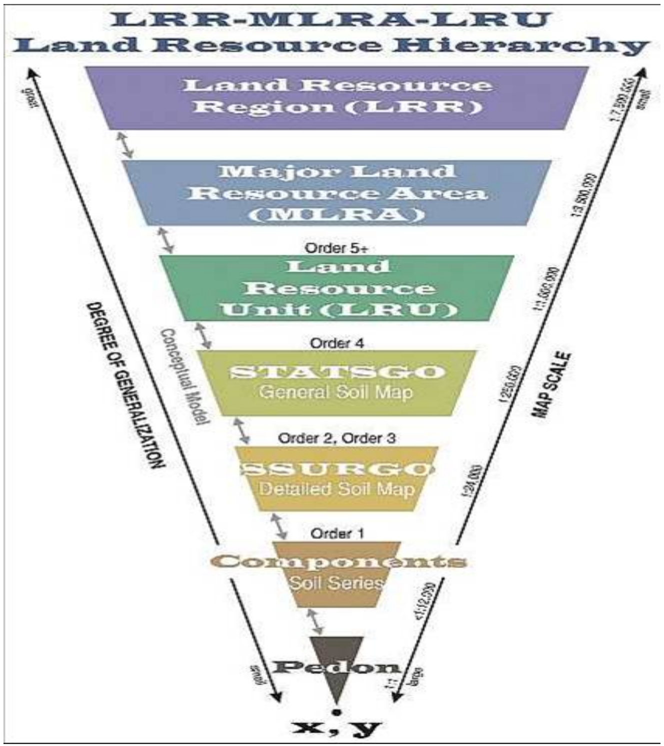


Figure 2: Status map for second-order soil survey (Source: Accessed from NRCS website on 23/06/08, http://soildatamart.nrcs.usda.gov/StatusMaps/SoilDataAvailabilityMap.pdf)

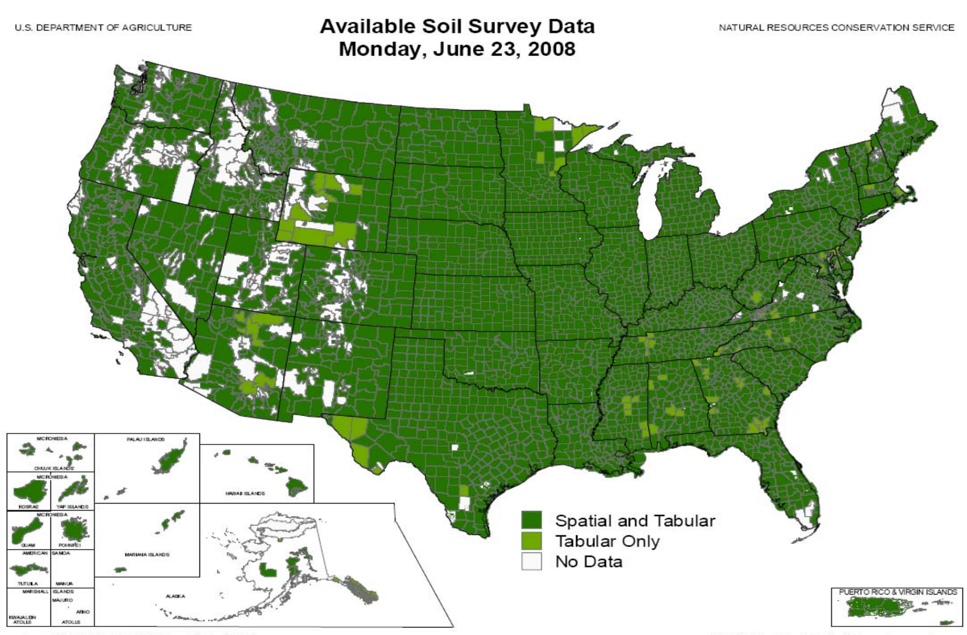


Figure 3: Welfare analysis in market equilibrium framework

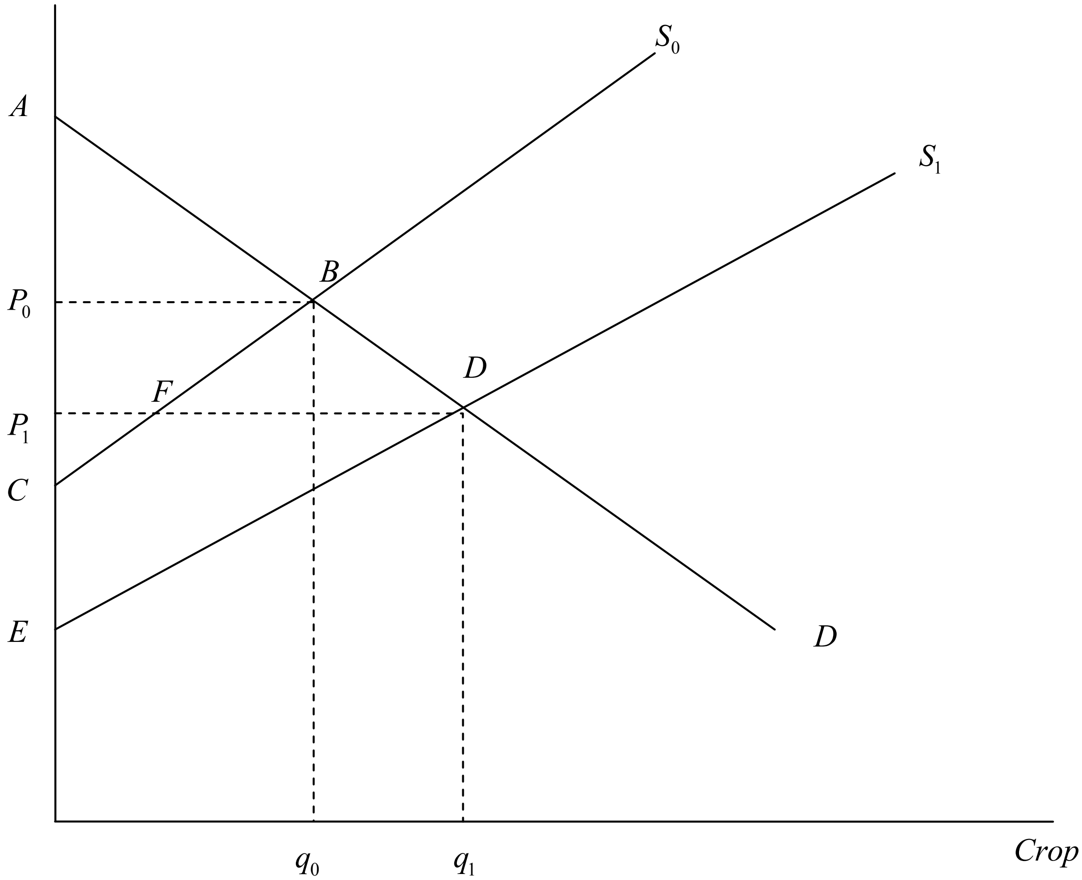


Figure 4: Selected counties

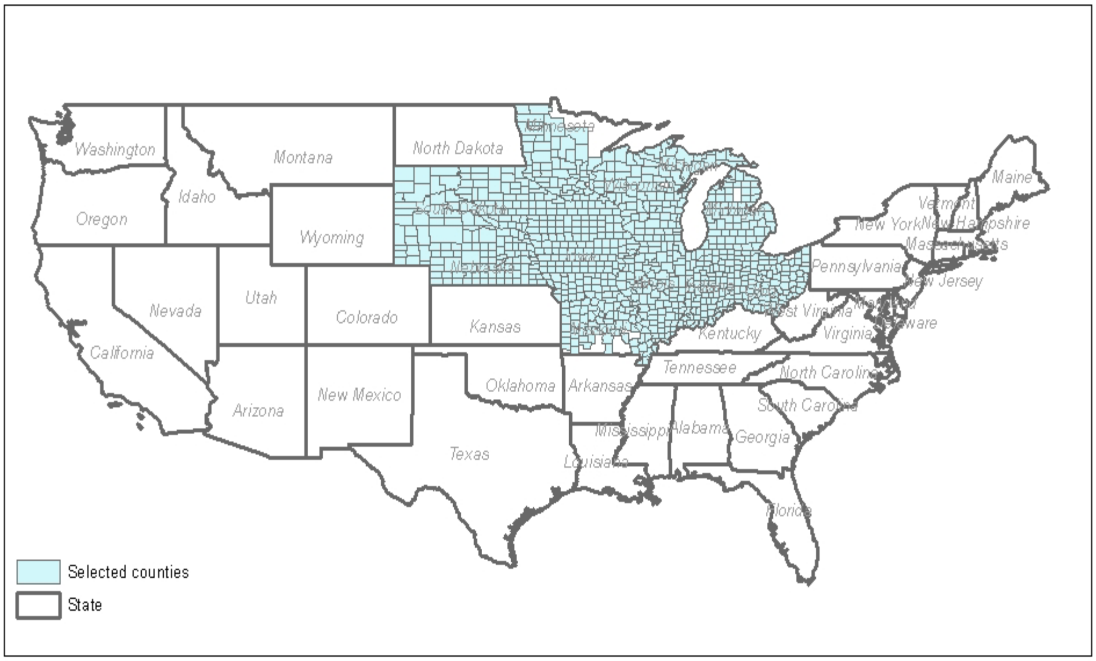


Figure 5: Proportion crop acres in corn

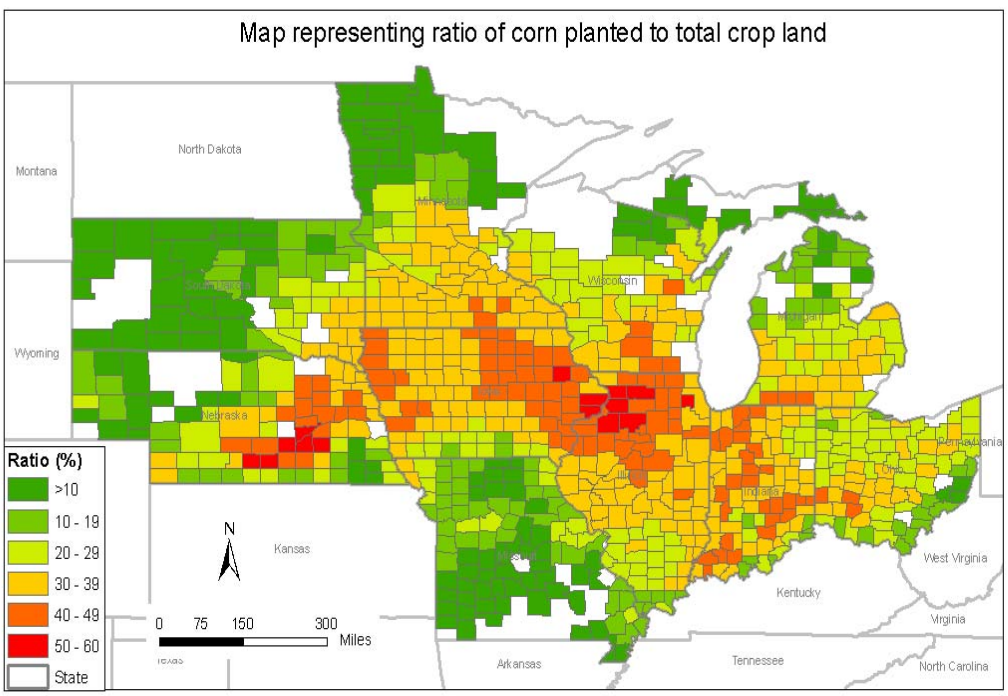


Figure 6: Location of NOAA stations with temperature records in the study area

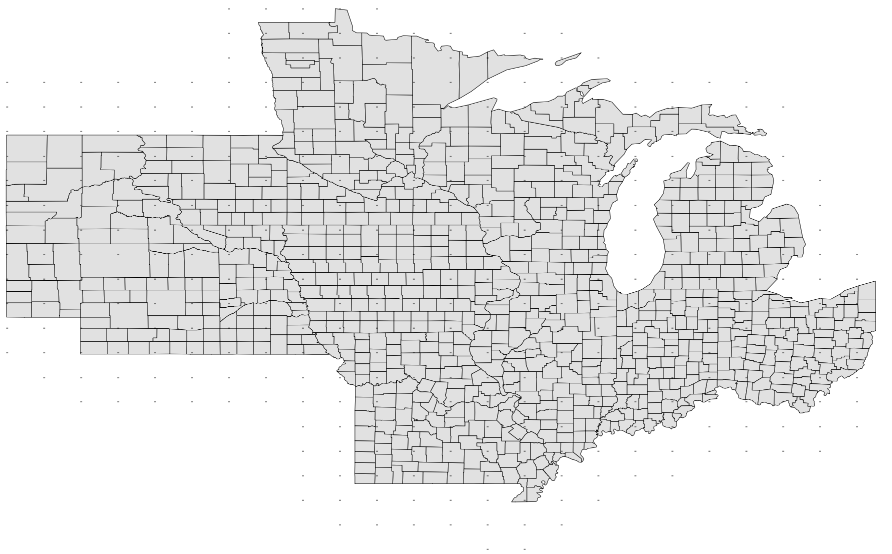


Figure 7: Mean and standard deviation of residual obtained from regression analysis

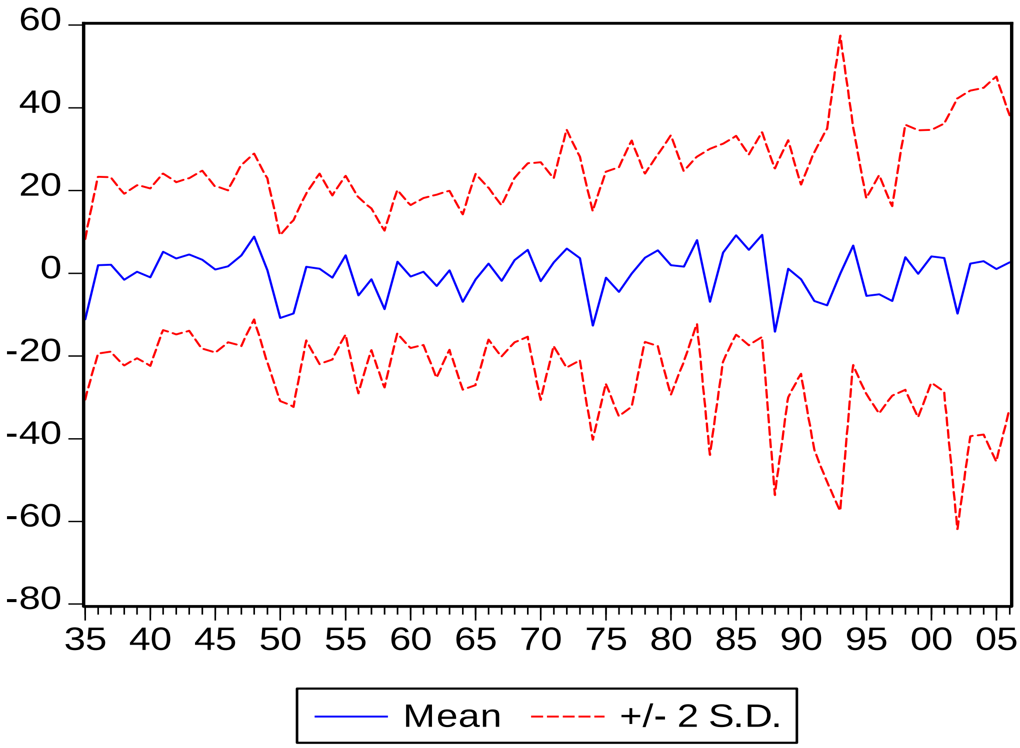


Figure 8: Moran's I index to test spatial autocorrelation of the residuals

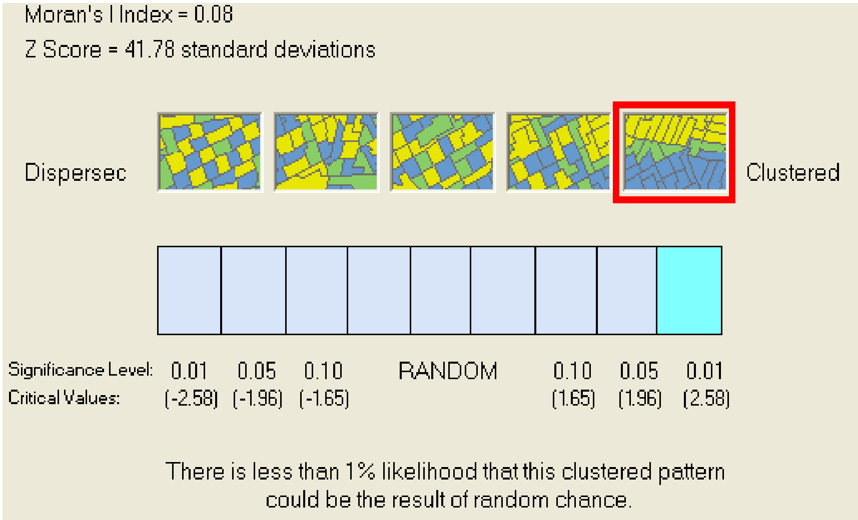


Figure 9: Soil Survey Certified date

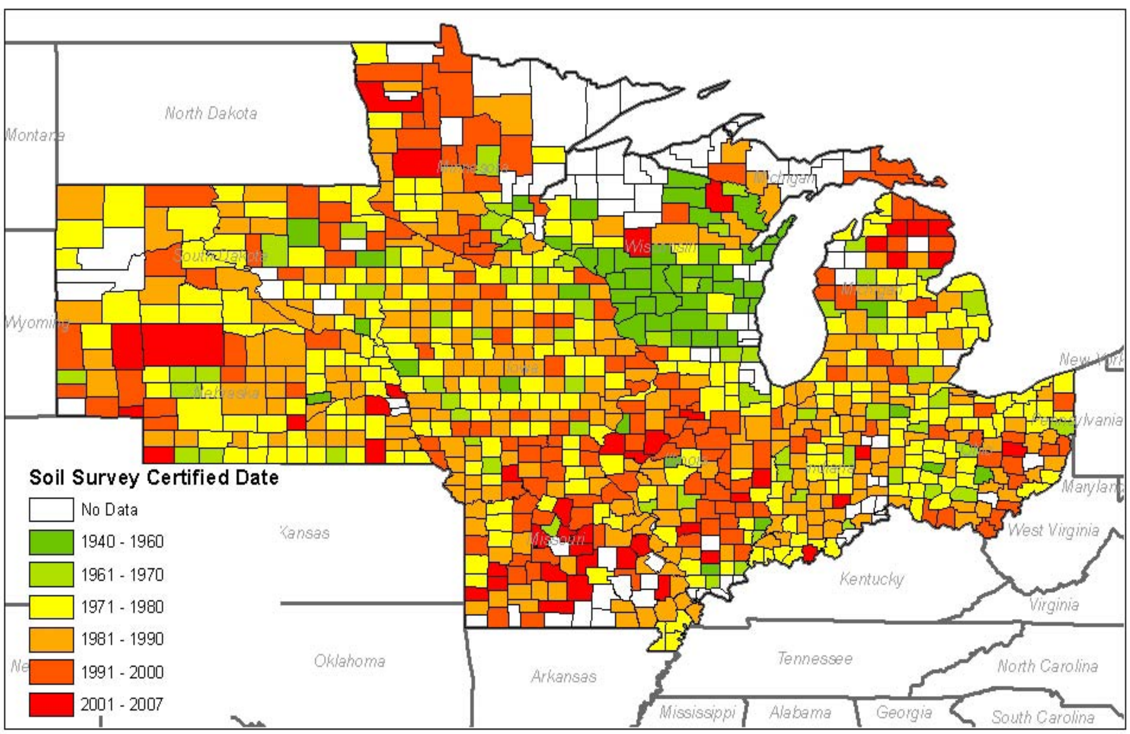


Table 1: Result from regression analysis using fixed effect model

Dependent Variable: YIELD
Method: Panel Least Squares
Sample: 1935 2006
Cross-sections included: 868
Total panel (unbalanced) observations: 60472

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	222.7466	2.965360	75.11621	0.0000
D_SSCORR1	2.503497	0.223478	11.20243	0.0000
TREND1	-2.399468	0.066863	-35.88630	0.0000
TREND2	0.177095	0.003640	48.65753	0.0000
TREND3	-0.003199	7.45E-05	-42.96191	0.0000
TREND4	2.00E-05	5.05E-07	39.58442	0.0000
D_1993	-28.74449	0.545301	-52.71305	0.0000
JUNE_MNT	0.418339	0.043835	9.543528	0.0000
JUNE_MXT	0.002410	0.038241	0.063010	0.9498
JUNE_PPT	-0.383492	0.116752	-3.284667	0.0010
JULY_MNT	1.386649	0.049551	27.98455	0.0000
JULY_MXT	-2.022975	0.042732	-47.34089	0.0000
JULY_PPT	3.228673	0.128350	25.15520	0.0000
AUG_MNT	-0.187795	0.044788	-4.193027	0.0000
AUG_MXT	-1.246933	0.040963	-30.44012	0.0000
AUG_PPT	0.459687	0.132028	3.481734	0.0005

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.870510	Mean dependent var	77.09762
Adjusted R-squared	0.868594	S.D. dependent var	40.36942
S.E. of regression	14.63391	Akaike info criterion	8.219055
Sum squared resid	12761070	Schwarz criterion	8.350616
Log likelihood	-247628.3	F-statistic	454.1889
Durbin-Watson stat	1.339142	Prob(F-statistic)	0.000000

Table 2: Comparison of weather coefficients with previous studies

	Our model		Garcia et al., 1987 (Illinois)		Kaylen et al.1992 (Corn Belt only among other regional zones)		Dixon et al. 1994 (Illinois)		Schroder et al. 1984	
	coef	prob	coef	prob	coef	prob	coef	prob	coef	prob
MAY_PPT			-3.00							
JUNE_MNT	0.418339	0.0000								
JUNE_MXT	0.002410	0.9498								
JUNE Temp			1.60				.794			
JUNE_PPT	-0.383492	0.0010								
JULY_MNT	1.386649	0.0000								
JULY_MXT	-2.022975	0.0000								
JULY Temp			-3.34				-1.23		-2.87	
JULY_PPT	3.228673	0.0000	1.03		2.76				5.31	
AUG_MNT	-0.187795	0.0000								
AUG_MXT	-1.246933	0.0000								
AUG Temp			-3.00		-2.75		-2.74			
AUG_PPT	0.459687	0.0005							1.96	

(Note: MNT = Minimum Temperature, MXT = Maximum Temperature and PPT= Precipitation)

References

Alexander, R.B. and R.A. Smith. 1990. County Level Estimates of Nitrogen and Phosphorus Fertilizer Use in the United States 1945 to 1985. US Geological Survey Open-File Report 90-130. Reston, VA.

Baltagi B. 1995. In: *Econometric Analysis of Panel Data*, Wiley, New York.

Beckett P. H. T. and P.A. Burrough. 1971. "The Relation between Cost and Utility in Soil Survey: IV. Comparison of the Utilities of Soil Maps Produced by Different Survey Procedures and to Different Scales". *J. of Soil Science*, 22(4): 466-480

Bie S. W., A. Ulph and P.H. T. Beckett. 1973. "Calculating the Economic Benefits of Soil Survey". *J. of Soil Science*, 24(4): 429-435

Bie S. W. and A. Ulph. 1972. "The Economic Value of Soil Survey Information". *J. Agricultural Economics*, 23: 285-297

Buller O. 1972. "Influence of Research and Policy on Crop Yields in Kansas". *Transactions of the Kansas Academy of Science*, 75(1): 20-28

Dent D. and A. Young. 1981. *Scale, Accuracy, Costs and Returns*. London: George Allen & Unwin, 278p.

Ditzler C.A., R.J. Engel, and R.J. Ahrens. 2003. "Soil Taxonomy and Soil Survey" In *Soil Classification: A Global Desk Reference.* Eswaran Hari, Ahrens Robert J, Rice Thomas J, and Stewart B. A. CRC Press, Washington, D.C.

Freeman A. M. III. 1993. *The Measurement of Environmental and Resource Values*. Washington, DC: Resources for the Future.

Garcia P., S.E. Offutt, M. Pinar and S.A. Changnon, 1987 "Corn Yield Behavior: Effects of Technological Advance and Weather Conditions". *J. Appl. Meteor.*, 26, 1092–1102

Greene W. 2003. *Econometric Analysis*. 5th edition. Prentice Hall: Englewood Cliffs.

Giasson E., C. van ES. and A. van Wambeke. 2000. "Assessing the Economic Value of Soil Information Using Decision Analysis Techniques". *Soil Science*, 165(12): 971-978

Griliches Z. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change". *Econometrica* 25: 501-522

Gujarati D. 2004. *Basic Econometrics*. Fourth Edition McGraw-Hill.

Houck J. P. and P. W. Gallagher. 1976. "The Price Responsiveness of U.S. Corn Yields". *Amer. J. Ag. Econ.,* 58:731-34

Hu Q. and G. Buyanovsky. 2003. "Climate Effect on Corn Yield in Missouri". *J. Appl. Meteor.*, 42: 1626-1635

Huff F. A. and J. C. Neill. 1982. “Effects of Natural Climatic Fluctuations on the Temporal and Spatial Variation in Crop Yields”. *J. Appl. Meteor*., 21: 540–550

Just R., D. L. Hueth and A. Schmitz. 2004. The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation. Edward Elgar, Cheltenham, UK, Northampton, MA, USA.

Kaufmann R.K. and S. Snell. 1997. “A Biophysical Model of Corn Yield: Integrating Physical and Economic Determinants”. *Amer. J. of Agri. Econ.*. 79(1):178-180

Kaylen M. S., J. W. Wade and D. B. Frank. 1992. "Stochastic Trend, Weather, and U.S. Corn Yield Variability." *Appl. Econ.* 24: 513-18

Klingebiel A.A. 1966. “Costs and Returns of Soil Survey.” *Soil Cons*., 32: 3-6

Macauley M. K. 2005. “The Value of Information: A Background Paper on Measuring the Contribution of Space-Derived Earth Science Data to National Resource Management.” Discussion Paper 05-26. Washington, DC: Resources for the Future.

McCall, John J., ed. 1982. *The Economics of Information and Uncertainty.* Chicago: University of Chicago Press.

McGee J. and L. Prusak. 1993. *Managing Information Strategically*. New York: John Wiley & Sons, Inc.

Menz K. M and P. Pardey. 1983. “Technology and U.S. Corn Yields: Plateaus and Price Responsiveness”. *Amer. J. of Agri. Econ.*. 65(3): 558-562

Mitchell, R. C. and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.

NRCS. http://soildatamart.nrcs.usda.gov/StatusMaps/SoilDataAvailabilityMap.pdf .(Accessed on June 23, 2008)

Offutt S. E., P. Garcia and M. Pinar. 1987. “Technological Advance, Weather, and Crop Yield Behavior”. *North Central J. of Agri. Econ.*. 9(1): 49-63

Rosenzweig, C., A. Iglesias, X.B. Yang, P. R. Epstein and E. Chivian. 2001. “Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests”. *Global Change & Human Health*. 2(2): 90-104 (http://chge.med.harvard.edu/about/faculty/journals/rosenzweig2001.pdf)

Samuelson J. R., M. Stelford, and R. J. Daniel. 2002. “The Importance and Value of Soil Information” 2002 ASAE Annual Meeting, Paper number 021093. St. Joseph, Michigan www.asabe.org

Schlenker W. and M. J. Roberts. 2006. “Nonlinear Effects of Weather on Corn Yields”. *Review of Agri. Econ.*. 28(3): 391-398

Schroder D., J.C. Headley and R. M. Finley. 1984. “The Contribution of Herbicides and Other Technologies to Corn Production in the Corn Belt Region, 1964 to 1979”. *North Central J. of Agri. Econ.*. 6(1): 95-104

Shen, S.S.P., P. Dzikowski, G. Li and D. Griffith. 2001. “Interpolation of 1961–97 Daily Temperature and Precipitation Data onto Alberta Polygons of Ecodistrict and Soil Landscapes of Canada”. *J. Appl. Meteor.* 40: 2162-2177

Simonson R.W. 1955. “Changing Place of Soils in Agricultural Production”. *The Scientific Monthly*. 81(4): 173-182

Smith D. W. 1998. “Soil Survey During Its Infancy of the Late 1890’s and Early 1900’s” National Cooperative Soil Survey Centennial Highlights. Papers for the 1998 Annual Conference of the Soil and Water Conservation Society, Presented July 1998, San Diego, California [Available online at ftp://ftp-fc.sc.egov.usda.gov/NSSC/NCSS/History/1998swcspapers.pdf]

Soil Survey Division Staff. 1993. Soil Survey Manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18. (http://soils.usda.gov/technical/manual/)

Swanson E.R. and J.C. Nyankori. 1979. “Influence of Weather and Technology on Corn and Soybean Yield Trends”. *Agricultural Meteorology.* 20: 327-342

Thomson L.M. 1975. “Weather Variability, Climatic Change, and Grain Production”. *Science.* 188: 535–541

USDA-NASS, cited 2007: Quick Stats: Agricultural Statistics Data Base. [Available online at http://www.nass.usda.gov/QuickStats/Create_County_All.jsp]

Western S. 1978. “Survey Quality and Survey Value”. In: S. Western, *Soil Survey Contracts and Quality Control*. 284f. Monographs (Soil Survey) - Oxford University.

Wooldridge J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*, MIT

Wooldridge, J.M. 2006. *Introductory Econometrics: A Modern Approach*, Third Edition, Thomson South-Western

Young A. 1973. “Soil survey procedures in land development planning”. *Geographical J.* 139: 53-64

Young R. A. 2005. *Determining the Economic Value of Water*: Concepts and Methods, Resources for the Future, Washington DC. ISBN 1891853988 (pbk.), 374 pp