

2014

MARKET MECHANISMS FOR VALUING PUBLIC GOODS

Julie Insignares-Santos
University of Rhode Island, jinsignares@my.uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Terms of Use

All rights reserved under copyright.

Recommended Citation

Insignares-Santos, Julie, "MARKET MECHANISMS FOR VALUING PUBLIC GOODS" (2014). *Open Access Dissertations*. Paper 207.
https://digitalcommons.uri.edu/oa_diss/207

This Dissertation is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

MARKET MECHANISMS FOR VALUING PUBLIC GOODS

BY

JULIE INSIGNARES-SANTOS

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ENVIRONMENTAL AND NATURAL RESOURCE ECONOMICS

UNIVERSITY OF RHODE ISLAND

2014

DOCTOR OF PHILOSOPHY DISSERTATION
OF
JULIE INSIGNARES-SANTOS

APPROVED:

Dissertation Committee:

Major Professor Emi Uchida

Corey Lang

Ruby R. Dholakia

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2014

ABSTRACT

Conventionally, inefficiencies in supplying socially optimal levels of natural amenities have been addressed by government intervention via taxes and subsidies. However, these measures may not result in the socially optimal level of provision because they are often influenced by collective action and may inadequately account for local tastes and preferences. This dissertation research addresses the ways in which private markets instead can be used to solve the dilemma of under-provision and over-exploitation of natural resources. The first manuscript examines a prominent ecolabeling program that provides consumers with information about responsible sourcing of seafood and investigates whether or not there is a demonstrated price premium in the retail market for ecolabeled seafood. The study specifically focuses on the commonly-voiced doubts regarding whether consumers are willing to pay a premium for the Marine Stewardship Council's (MSC) label. The analysis utilizes scanner data on frozen pollock products from supermarkets in the metropolitan area of London, UK across a sixty-five week panel. I use a hedonic model to control for brand, package size, product type and form. I find evidence that consumers within the sampled area are paying a 14% premium for MSC-certified pollock products. The second and third manuscripts address the possibility of creating a direct market for natural amenities in which consumers can contract with suppliers or custodians of the resource. However, complications arise when eliciting preferences for natural amenities because they are often public goods and thus are non-rival and non-excludable, i.e. scenic views, clean air and drinking water, etc. Willingness to pay for public goods are difficult to measure because individuals have incentives to hide their true values for the good offered for exchange. This research deals with the two most prominent sources of this bias: incentives to free-ride on others' contributions and the tendency of respondents to overstate their values when no monetary consequence exists. The second manuscript explores how well laboratory-tested public goods elicitation mechanisms that mitigate

free-riding perform in the field. I employ a mixed logit model in willingness-to-pay space to estimate individual-specific willingness to pay for protecting grassland nesting bird habitat on farmland and compare some of the most promising elicitation mechanisms in their ability to produce results that yield the highest valuations. I also estimate individual-specific measures of the scale of the error variance and test which elicitation mechanisms are the least noisy. I find evidence that providing a familiar reference mechanism induces behavior more in line with laboratory results, but that otherwise, individuals tend to ignore the elicitation mechanisms in the field. The third manuscript addresses a continuing debate [1, 2] in the literature regarding whether hypothetical choice experiment surveys can accurately reflect revealed preferences in a market for public goods. This issue is particularly important in studies involving valuation of public goods and common-pool resources because non-rival non-exclusive goods imply a disconnect between what is paid for and what can be consumed. I utilize a latent class model of attribute non-attendance to identify individuals who are more or less likely to accurately state their willingness-to-pay for habitat preservation for grassland nesting birds on farmland. I find that hypothetical bias differs based on the strategy employed in the stated preference experiment and that individuals who are estimated to fully attend to all attributes are more likely to generate reliable preference estimates. Overall, this research provides guidance on market approaches for enhancing environmental public goods provision.

ACKNOWLEDGMENTS

With a heart brimming with love, I would like to express my most sincere gratitude to all of the people who have helped me along this arduous journey. Most importantly, to my two great major advisers, Dr. Emi Uchida and Dr. Cathy Roheim, whose efforts and encouragement were exceptional, indispensable, and comforting at all the right times. To Emi for providing her expertise and time when I really needed it, for accommodating my last minute conference presentation jitters, and for being such an exceptional source of ideas and resources. And to Cathy, who supported me during one of the most difficult moments in my life, offered her compassion and has been a model of strength and perseverance to me. I am truly blessed to have had the support of two such brilliant, inspiring women.

I would also like to thank the members of my dissertation committee: Dr. Ruby Dholakia, Dr. Corey Lang, Dr. Hiro Uchida, and Dr. Gavino Puggioni for their excellent input and advice.

I am forever grateful to my mother, Irene, who always speaks the truth, even if you're trying to avoid it, and always with love and kindness. She has given me strength to do the same. And to my father, who according to all Chinese fortune cookies, is the greatest most handsome man alive, and that's pretty much the way I see him, too. I am so honored to have such giving and loving parents. I'd also like to thank my brother, Mikey. I'm not sure if there is another human being who knows me better. Although we'll never agree on whose music is better, I am so comforted by the thought that we always have each other's back.

I'd like to humbly thank my son, Jake, for his patience when mommy was at work and missing out on play time and my loving husband, for helping out even when he'd rather veg out.

I am also deeply indebted to dear friends and colleagues: Tracy Hammond, Erika Kayden, Dana Seaton, Gilbert Magbag, Hiroki Wakamatsu, Sudha Kannan, Erika Myers, Sha Sha, Todd Szahun, Suzanne Mathieu and Linda Cortinas.

And, last but not least, thanks to my dog, Milo, a faithful friend, who has sacrificed birthday hikes and beachcombs for the good of the cause.

DEDICATION

You have brains in your head.

You have feet in your shoes.

You can steer yourself any direction you choose.

-Dr. Seuss

I dedicate this work to my son, Jake, the light of my life.

PREFACE

This dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in Environmental and Natural Resource Economics is in the manuscript style format. The dissertation is composed of three manuscripts.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
DEDICATION	vi
PREFACE	vii
TABLE OF CONTENTS	viii
MANUSCRIPT	
1 The Elusive Price Premium for Ecolabeled Products: Evidence from Seafood in the UK Market	1
1.1 Introduction	1
1.2 Background	3
1.3 Data	6
1.4 Model Specification and Estimation	9
1.5 Results	11
1.6 Conclusions	14
2 Factors that influence response reliability in survey data: Addressing challenges of employing incentive-compatible mechanisms outside the lab	20
2.1 Introduction	20
2.2 The Experiment	24
2.2.1 The Elicitation Mechanisms	26
2.3 The WTP-space approach	28
2.3.1 Model Specification	33

		Page
2.4	The Model and Estimation	35
2.4.1	The Mechanics of HB Estimation	35
2.5	Results	39
2.5.1	Distributions of Willingness to Pay	40
2.5.2	Root Likelihood	41
2.5.3	Factors that Influence Scale	43
2.5.4	Scale and The Rules of the Game	45
2.6	Summary and Conclusions	48
3	Predictive Validity in Value Assessment For the Provision of Farmland Amenities: A Disaggregated Approach	61
3.1	Introduction	61
3.2	Constructing a local market for wildlife preservation	64
3.3	Methods	69
3.3.1	The LCL Model	70
3.3.2	Constructing Individual-level WTP estimates	75
3.3.3	Analyzing Market Responses	76
3.4	Results	77
3.4.1	Descriptive Statistics and Class Allocation	77
3.4.2	Market Experiment Outcomes	80
3.4.3	Measuring Hypothetical Bias and SP undervaluation	83
3.5	Summary and Conclusions	84
APPENDIX		
A Elicitation Mechanisms Employed to Mitigate Free-Riding		105

		Page
A.1	The Stated Preference Mechanism Descriptions	105
A.2	Market Experiment Mechanism Descriptions	109
B	Convergence Plots and Descriptive Statistics for the Bayesian Analysis: Manuscript 2	111
B.1	The Models	111
B.2	Convergence Indicators	112
B.2.1	Autocorrelations and Lags	113
B.2.2	Trace Plots	114
B.2.3	Density Plots	117
B.2.4	Log Plots	120
	LIST OF REFERENCES	123

Introduction

Societies rely on private markets to allocate scarce resources from least-cost providers to those who most value goods and services. Because well-defined property rights are required to ensure a healthy, efficient market, achieving the socially optimal level of output of public goods and common-pool resources (CPR) through private market exchange alone is impossible. For instance, several ecosystem services provided by nature, such as wildlife habitat, open-space and groundwater aquifer recharge certainly have value to local communities and society in general. Unfortunately, land owners must choose between land uses that yield private market value, such as high yield farming practices or industrial construction, and those that do not have direct or immediate market benefits.

The typical remedy for under-provision of public goods and over-exploitation of common-pool resources involves government intervention, often in the form of levied taxes, gear restrictions, cap-and-trade programs, or other regulations. Most common government actions are aimed at producers and entail supply-side policy instruments. However, these measures may not result in the socially optimal level of provision because they are often influenced by collective action by special interest groups and may inadequately account for local tastes and preferences. The overall goal of this dissertation is to evaluate ways in which market incentives can be provided to producers of market goods in order to incorporate practices that are consistent with improving ecosystem health and mitigating the externalities involved in manufacturing private goods. Specifically, I examine the demand-side instruments that can drive these incentives. Ultimately, this research attempts to shed light on the promising opportunities and significant challenges that exist for market-based approaches for bolstering public goods provision and environmental sustainability.

To achieve this goal, I explore the effectiveness of two strategies to provide private incentives for producing public and CPR goods. The first manuscript focuses on eco-labeling as a tool to empower consumers to vote for environmentally friendly production practices. The second and third manuscripts provide an analysis of a novel experiment in which a local market-like process for wildlife-habitat protection was designed and implemented.

Ecolabels, such as organic labeling, energy-efficiency labels, the Forest Stewardship Council's (FSC) ecolabel on wood products and the Marine Stewardship Council's (MSC) label on seafood are gaining in popularity as consumers become increasingly aware of the impacts of their purchases on the natural environment around them [3]. These labeling programs allow consumers to differentiate between items not solely on their physical characteristics but also on unobservable attributes of the production process. But, the effectiveness of the label depends on consumer awareness of the label and trust in its claims. In addition, there must be some demonstrable demand for the environmental attributes touted by the label. Many have voiced skepticism regarding whether ecolabels are an effective tool for encouraging responsible resource management. With regard to seafood products, several studies have presented evidence that consumers are hypothetically willing to pay a premium for sustainably harvested seafood [4, 5, 3, 6, 7]. However, little is known about whether these intentions extend beyond intentions and to what extent ecolabels on seafood yield quantifiable market benefits. The existence of a price premium being realized for labelled products achieves two goals. First, it is a step toward quelling doubts about the effectiveness of ecolabels on seafood and, second, it provides some information to fisheries considering certification as to the potential benefits of pursuing certification.

In this first manuscript, sales of frozen pollock products from supermarket scanner data in the London, UK area are analyzed for evidence of a price premium for the

MSC's eco-label on sustainably harvested pollock. The choice of frozen pollock products is based on the fact that the Alaskan pollock fishery was certified against the Marine Stewardship Council's requirements fairly early in the history of the labeling program. It is thus more likely that consumers are familiar with the label. I use an hedonic model to control for other physical attributes of the products, such as brand, process form, and size. The results indicate that, on average, the MSC label generates a 14% price premium in the market for frozen pollock products in the sampled area. This research provides strong evidence that market benefits are being realized for sustainably harvested seafood.

Although eco-labeling programs serve as a vehicle through which consumers can express their preferences for the treatment of the environment, a direct market approach places the environmental resource at the center of the economic exchange. Constructing a market approach for the resource itself opens a path from consumer demand to producers' supply without supply chain or other considerations. This advantage, however, comes with its own set of costs and complications. Beginning in 2006, researchers at the University of Rhode Island orchestrated a novel field experiment designed to elicit direct values for the private provision of an agricultural ecosystem service: that of grassland nesting habitat protection for the Bobolink bird species. Jamestown is a small community of 4500 residents inhabiting approximately 2800 households. The residents have a history of support for community-wide conservation projects, particularly for farmland and open-space. The experiment involved the residents and farmers of Jamestown, Rhode Island and tested several elicitation mechanisms that were shown to mitigate free-riding in the lab. Free-riding, a result of treating a public, non-excludable good as a marketable good, has been found to yield inefficient outcomes in public goods experiments. It is well-known and documented, particularly in philanthropic endeavors, that an individual has some incentive to free-

ride on the contributions of others if she believes that provision is not contingent on contribution of her full value [8, 9, 10].

In the second manuscript, I investigate the efficacy of different elicitation mechanisms to mitigate free-riding in the field. There has been a significant effort within the experimental economics scholarship in constructing mechanisms that will create incentives for players to reveal their true values in public goods experiments. Some of the more effective mechanisms involve some combination of provision points, money-back guarantees, and proportional rebate of excess funds. The pivotal mechanism particularly, which collects an individual's bid only if her contribution is pivotal to the provision of the good, has been shown to be, under certain circumstances, incentive compatible in the lab. However, few of these mechanisms have been tested in field experiments [8, 10]. Some authors have suggested that mechanisms that prove effective in the lab may not perform as expected in the field because they may be too complex or unfamiliar [11, 12]. Field experiment participants do not have the option to ask questions if an instruction is not clear and the researcher has little control over how the information is processed.

To investigate these dynamics, the second manuscript utilizes a stated preference choice experiment survey that was designed to elicit values for a bundle of environmental amenities associated with preserving grassland nesting habitat on Jamestown's farms. Choice experiment data have been used to measure preferences for environmental amenities such as: forest landscape features [13], improvements in river ecology and aesthetics [14], recreational demand for different aspects of rock climbs in Scotland [15], nature conservation programs in Finland [16], passive use values for Caribou preservation [17], aspects of environmental projects managed by the World Wildlife Foundation [18], features of competing policy measures relating to soil erosion in Southern Spain [19] to name a few. Based on random utility theory, the

choice experiment design lists several options as packages of features or attributes and asks participants to choose the most desirable option based on the listed attributes. The Jamestown choice experiment elicited preferences for restoring fallow farmland to active cultivation, scenic views, and an expert-led bird walk in addition to protected habitat for nesting birds. The choice experiment format is widely accepted as the highest standard survey technique for generating reliable valuation in hypothetical settings. Prior research in valuation of public goods employed simple voluntary contribution mechanisms. The Jamestown experiment administered several different public goods elicitation mechanisms to participants in the field including: the pivotal mechanism based on the Clarke tax, uniform price auction, and the proportional rebate mechanism.

I focus on differences in choice consistency and willingness to pay estimates from the various elicitation mechanisms. Choice consistency acts as a proxy for choice task complexity and cognitive burden. Several researchers have examined heteroskedasticity in error variance in relation to choice experiment design dimensions such as the number of choices in the choice set or the number of attributes per alternative as well as learning and fatigue effects [20, 21, 22, 23, 24]. Along these lines, I examine choice consistency by elicitation mechanism in order to address the contention that demand revealing mechanisms do not perform as expected in the field because they are too obscure or complex. I augment the analysis by comparing willingness to pay across mechanism treatments. This permits an assessment of the ability of the elicitation mechanisms to mitigate free-riding behavior.

I implement a random parameter mixed logit model in willingness-to-pay space to retrieve estimates of scale for each respondent. I employ a hierarchical Bayes model of population outcomes as hyper-parameters for the individual-level estimates. The individual-specific scale estimates are then analyzed using linear regression techniques

to analyze the determinants of scale heterogeneity paying particular attention to the alternative public good mechanism treatments.

The results suggest that demand-revealing public goods mechanisms are capable of performing in the field in a manner similar to results obtained via laboratory public goods experiments. However, there is some evidence that a familiar base mechanism must be applied to draw respondents' attention to the new mechanism. This research provides important insights into the design of efficient market mechanisms for generating funds for public goods projects.

Yet, there is a second drawback to the market approach—that of private value construction for a good that has previously been publicly consumed. Individuals conventionally contribute funds to environmental causes via charities and other philanthropic efforts. Often, they are asked for their contribution under the presumption that what is contributed goes toward a general goal: providing public radio, saving rainforests, etc. They need only be concerned that the charity, being a reputable, legitimate institution and having expertise with the problem at hand, will put the donation to its best use.

In the choice experiment, the participant is given precise information about the environmental practice, amenity or policy she is expected to value. Several natural (dis)amenities are listed in combination with varying levels, and the relative marginal values (costs) are inferred from the individual's choices. However, there is no frame of reference within this context. Asking an individual what she might pay to ensure that some number of wildlife mortalities will be spared, without reference to the scope that the effort might have on the health of the wildlife stock in the future may prove problematic. One might be inclined to simplify the choice to make it more like a charitable donation setting, either by focusing on the cost of the bundle of attributes or by choosing the attribute of the bundle that most matters to her.

As choice modeling research has shifted in recent history to focus on the sources of heterogeneity in responses of individuals to choice experiment scenarios, one line of study recognizes that individual behavior may be better described by alternative representations of utility functions that accommodate attribute processing strategies. The most commonly investigated variation on the additive linear utility function is attribute non-attendance. Cameron and DeShazo [25] have pointed out that individuals make trade-offs between costs of further information processing and expected marginal benefits of attributes in choice experiments. This leads respondents to attend to some attributes and overlook others. Hensher and Rose [26], Hess and Hensher [27], Hensher et al. [28] and other research have highlighted the importance of accommodating information processing strategies for behavioral outcomes of choice experiments.

The third manuscript addresses the possibility that attribute processing rules can be used as a means of identifying hypothetical bias in stated preference data. I compare responses to the hypothetical, stated choice experiment implemented in 2006 with the respondents actions in an implementation of the market-like process. In 2007 and 2008, researchers from the University of Rhode Island approached the farmers of Jamestown to assess the costs of delaying harvest on their fields long enough to ensure that young birds have had adequate time to fledge and avoid mortality due to harvesting activities. The community was subsequently solicited to contribute to contracts to provide compensation to the farmers. Individuals provided offers by returning a survey with a monetary commitment in the form of a check or credit card authorization.

I use a latent class model that accommodates attribute processing rules such as yea-saying, attribute non-attendance, and lexicographic preferences to predict participation and valuation in the field market experiments in 2007 and 2008. The

results suggest that accommodating various information processing strategies allows the researcher to make predictions about behavior in settings that have monetary consequences with greater accuracy. The main finding from this research is that individuals who expend the cognitive effort to fully attend to all attributes in the hypothetical realm also exhibit highest valuations in revealed preference scenarios. In addition, I find that yea-saying is a significant challenge in stated preference data and hypothetical bias for this strategy of response was quite high.

The methods and results outlined in this research have broad implications for designing markets for ecosystem services and validating valuation outcomes from stated and revealed preference studies. While more work is left to be done, this research makes a significant step toward understanding how market mechanisms can be used to enhance provision of public goods from the environment.

**The Elusive Price Premium for Ecolabeled Products: Evidence from
Seafood in the UK Market**

Cathy A. Roheim¹, Frank Asche², and Julie Insignares Santos³

Published in Journal of Agricultural Economics

May, 2011

Volume 63, Issue 3, pages 655-688

¹Professor, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI, 02881. Email: crw@uri.edu.

²Professor, Department of Industrial Economics, Risk Management and Planning, University of Stavanger, N-4036 Stavanger, Norway.

³PhD Candidate, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881. Lead author for Data, Results and Conclusion section. Co-authored Introduction, Background, and Model Specification. Performed all data preparation and analyses.

MANUSCRIPT 1

The Elusive Price Premium for Ecolabeled Products: Evidence from Seafood in the UK Market

1.1 Introduction

Governance of common pool resources, such as fisheries and publicly-owned forests, often fails to correct for over-exploitation of the resources [29]. In some cases, poor management may evolve from a close relationship between the managers and the industry being managed. As a result, decisions regarding what is best for the resource are replaced by decisions regarding what is best for those utilizing the resource. To allow consumers a voice, certification programs for sustainably-managed resources and ecolabeled products derived from those resources have been introduced.

The goal of ecolabeling programs is to create market-based incentives for better management of the environment. Ecolabels provide otherwise unobservable information to the consumer about the environmental attributes of goods, allowing consumers to differentiate between products carrying the labels from those which do not [30]. If consumers value the environmental attributes of the products conveyed by the ecolabel, they will shift their demand toward the ecolabeled products and away from the non-labeled products. This in turn may create a price premium for ecolabeled products over non-labeled products, thereby creating a market incentive for producers to supply those environmental attributes [31, 32].

Ecolabeling has become an increasingly important tool in the promotion of sustainable forestry and seafood products around the world [33, 34, 35]. In relation to seafood, the approach has created significant attention in markets since the first capture fishery was certified as sustainable against the Marine Stewardship Councils

(MSC) standards in 2000¹. Those who sell products from fisheries which are MSC certified may purchase licenses for the right to place the MSC ecolabel on affiliated products, signaling to consumers that the product was produced from a sustainable fishery. Although there are now competing labels, the MSC is the leading label in terms of the number of fisheries certified, volume of edible seafood certified, and logo presence in the global marketplace [36].

Whether the consumer is paying a price premium for ecolabeled products is one indicator of the market effectiveness of ecolabeling schemes. Yet there is a surprising lack of literature demonstrating the existence of price premiums for either seafood or wood products. A number of studies have found evidence that consumers indicate a preference for ecolabeled seafood, including Wessells, Johnston and Donath [4], Johnston et al. [5], Jaffry et al. [3], Johnston and Roheim [6], Brcard et al. [7], and Salladar et al. [37], and for ecolabeled wood products, including Ozanne and Vlosky [38, 39], Forsyth, Haley and Kozak [40], Teisl et al.[33] and Aguilar and Vlosky[41]. However, this literature is based upon contingent valuation survey data and show only that consumers state a preference for ecolabeled products under certain conditions. Determining the existence of actual price premiums in the market for ecolabeled products is important to address the expressions of skepticism by policy makers and others regarding the effectiveness of ecolabeling as a tool to create more effective management [42, 43]. Such skepticism exists due to the lack of rigorous evidence that consumer preferences have transformed into actual price premiums for the certified fisheries.

The purpose of this paper is to investigate whether or not there is a demonstrated price premium in the retail market for ecolabeled seafood. A demonstrated price premium paid at the retail level may, depending upon the price transmission mechanism,

¹Similarly, the Forestry Stewardship Council (FSC) is a leading sustainable forestry certification body

indicate compensation at the fish production level. Because the focus of the analysis is upon the marginal value of the ecolabel, in other words the price premium, we follow Rosen [44] and use a hedonic price model. Our empirical analysis is applied to scanner data on frozen processed pollock products in the London metropolitan area. Scanner data are increasingly being used in price analysis, including hedonic analysis of price premiums for such labels as fair trade coffee [45] and organic agricultural products [46, 47, 48].

The paper proceeds as follows. The next section presents a brief background on the MSC program and seafood ecolabeling, as well as a discussion of the rationale behind the research focus on a retail market in the UK and MSC-labeled pollock. This is followed by a description of the data used, and a discussion of the model specification and estimation procedure used to measure the price premium. Results and implications from the model are discussed next, followed by the concluding remarks.

1.2 Background

The MSCs fishery certification program and seafood ecolabel recognize and reward sustainable fishing. The earliest fisheries certified were Alaskan salmon, Western Australian rock lobster and Thames River herring in 2000. In 2005 when the Alaska pollock (*Theragra chalcogramma*) fisheries in Alaskas Gulf of Alaska and Bering Sea were certified, the number of certified fisheries was only thirteen. In contrast, the number of certified fisheries as of January 2011 was 102. The number of fisheries in assessment for certification as of January 2011 was 132. As the number of fisheries in the program has grown, so has the market for MSC-labeled products. There are more than 5,000 MSC-labeled products on sale globally in over 66 countries with a retail value of over US \$2 billion annually [49].

Alaska pollock represents one of the largest fisheries in the world, and, in spite of the growth of the MSC program, the largest proportion of volume among the certified

fisheries within the MSC program. This fishery has an average annual historic catch of approximately one million metric tons during the past 30 years, with catch levels set by the U.S. federal fisheries management [50]. The primary markets for Alaska pollock are North America, Europe, and Japan.

Whether the consumer is paying a price premium for ecolabeled products is one indicator of the market effectiveness of ecolabeling schemes. [31] provide a theoretical framework showing that price premiums play a critical role in providing market-based incentives to the fishing sector for improving or maintain sustainable fishing practices. Wessells, Johnston and Donath [4], Johnston et al.[5], Jaffry et al. [3], Johnston and Roheim [6] and Brcard et al. [7] have shown empirically that some consumers prefer ecolabeled seafood products over non-labeled, and a statistically significant proportion indicate a hypothetical willingness to pay a positive premium, while Ozanne and Vlosky [38, 39], Teisl et al. [33] and Aguilar and Vlosky [41] show the same for ecolabeled wood products. However, these studies have in common that they use hypothetical data. The studies do not provide an estimated premium, and actual price premiums paid in the market, if any, are yet to be determined.

In spite of the results of these studies, Gulbrandsen [51] argues that most markets for ecolabeled forestry and fisheries products have been created as a result of pressure by environmental groups on consumer-facing corporations, rather than resulting from consumer demand. O'Brien and Teisl [52] go so far as to say that ecolabels are ineffective due to a lack of marketing, leading to a lack of consumer awareness of ecolabels in forest certification. Roheim [35] concurs, positing that the market for ecolabeled seafood is driven less by consumer demand than by corporate decisions to source certified sustainable seafood for a variety of reasons, including risk reduction. On the other hand, according to Sedjo and Swallow [32], even though consumers may express interest in purchasing an ecolabeled product, this does not suffice as evidence

that a price premium will manifest in the market. These academic discussions provide an impetus for research to document whether or not actual price premiums are being paid.

However, the most basic reason to determine the existence, if any, of price premiums in the market is to assist in evaluation of effectiveness of the ecolabeling as a market-based incentive. Producers of certified products and those contemplating assessment for certification are increasingly demanding proof of market benefits to justify the costs of the assessment process and of practicing sustainable fishing. For instance, according to Washington [43] and Roheim [53], the costs of obtaining MSC certification can range from \$10,000 for small and simple fisheries to \$500,000 for large and complex fisheries like the U.S. pollock fishery. Maintaining certification creates additional costs. However, costs of certification are only a fraction of the costs of transitioning to a sustainable fishery from a fishery which previously did not meet the conditions for sustainability. A sustainable fishery requires investment in appropriate fisheries management, practices, and capital. None are costless. Before entering the certification assessment process, fisheries must perceive that market benefits will be enough to offset these costs. Market benefits may not simply be a price premium as such, but may also include improved market access to premium markets, expanded market share in existing markets, and greater ability to favorably position oneself in the market with competitors [35].

Downstream in the supply chain, those who must have chain of custody certification also seek quantitative proof of the presence or absence of market benefits [35]. These firms have invested significantly in the program by sourcing certified product and paying license fees to place the logo on the products. A positive return is expected on this investment.

There are several reasons why analysis of premiums on MSC-labeled pollock products in the UK market is an ideal starting point from which to evaluate market benefits of fisheries certification. The UK market established itself early in sustainable seafood sourcing. This development was to a large extent driven by seafood processors and supermarket chains with strong commitments to the MSC and sustainable seafood sourcing [35, 54]. We focus on Alaska pollock products since the Alaska pollock fisheries were certified relatively early (in 2005). As such, enough time has elapsed to ensure that a significant number of MSC-labeled pollock products are in the market. Furthermore, UK supermarkets carry a number of frozen processed seafood products specifically-labeled as pollock (not whitefish) both with and without the MSC label. Finally, one of the essential conditions of a successful ecolabeling scheme is familiarity with and exposure to the label. Consumers in the UK are most likely to be aware of the label because of greater marketing of the MSC program within the UK through the Fish and Kids program, events hosted by Prince Charles, and other marketing efforts (www.msc.org). Jaffry et al. [3] confirms consumer interest in ecolabeled seafood in the UK. For these reasons, the focus of this paper is on the UK retail market for pollock.

1.3 Data

To assess the existence and size of retail price premiums for MSC-labeled Alaska pollock, scanner data were used in the estimation of a hedonic model. Scanner data from retail sales of products became widely available in the 1980s and are based on stock-keeping units (SKU) or bar code scanning at the supermarket check-out counters. For this project, Information Resources Inc. (IRI) Infoscan data was purchased for the London metropolitan market area. Infoscan data measures specific product flow through supermarkets. The data is aggregated over a sample of several hundred supermarkets. It differs from scanner data based upon household panels in that it

does not collect data on individual consumers at individual supermarkets. Several studies have indicated that, under certain circumstances, aggregate level data may produce biased estimates [55, 56, 57]. Theil [57], however, points out that this is a problem in linear models only if stores have heterogeneous responses to marketing actions. We have no information on the ways in which individual supermarkets respond to the marketing actions of manufacturers of frozen processed pollock products. However, it is not unreasonable in this case to assume homogeneous response to such actions across supermarkets. In addition, Cotterill [55] argues that even with aggregation, the market data contained in IRI Infoscan data allows for a rich set of possible empirical insights. For example, several researchers have used Infoscan data to investigate the effect of other types of labels. Lusk [58] provides an evaluation of demand for cage free, organic and conventional eggs, and Cotterill, Putsis and Dhar [59] and Cotterill and Putsis [60] in analyses of the competition between national and private labels of breakfast cereals and carbonated beverages.

The Infoscan database provides volume and dollar sales by SKU for over 400 frozen processed seafood products aggregated across supermarkets in the London metropolitan area on a weekly basis for 65 weeks, from February 24, 2007 to May 17, 2008. Unit prices for each product are averages derived from total sales and volumes. The focus is on processed food products since these are the only products for which SKUs are consistent across all supermarkets. SKUs on fresh products sold are specific to stores or supermarket chains, therefore fresh fish could not be included in the analysis.

For each pollock SKU, information provided includes: brand, species, product type (such as breaded, battered, natural smoked), product form (such as fillets, fish fingers, and kids fish - products in various fun shapes), and package size. Brands include national labels such as Youngs and Birds Eye or private labels. Private labels are labels associated with individual supermarket chains. The Infoscan database does

not specifically identify the label, thus these products are simply identified as private label products.

A total of 24 pollock products are included in the analysis. These products are similar in product form (fillets, fish fingers, kids fish). Pollock products which are highly value added, such as ocean pies or in which vegetables are added, are excluded. The Infoscan database does not contain information on which products carry the MSC label. Working with the logo licensing manager at the MSC, viewing the products on the websites of the producers and supermarket chains, and contacting the producers directly for affirmation, twelve of those products were identified as displaying the MSC ecolabel. Each national brand marketed both ecolabeled products and non-labeled products, indicating that these brands have differentiated individual product lines. Of the nine Youngs products, seven appear with the MSC label. Of the eleven Birds Eye products, five have the MSC label, while none of the four private label products have the MSC label. There were three kids fish products, ten fillet products, and eleven finger products. Package sizes varied from a low of 240 grams to a high of 1080 grams.

In all, a panel dataset of 1,137 observations were included in the analysis, one observation for each week when the twenty-four products were sold. During the sixty-five week time period, none of the products appear in the market for the entire period. Some products are introduced, and other products are withdrawn, from the market during the observation period. Given negligible inflation during the short time period, nominal prices are used.

The appearance of an MSC-labeled product on stores shelves, with or without a price premium, does not guarantee that consumers will purchase the product. Pollock historically has been sold as a generic whitefish in the UK market. After certification it has been specifically identified as pollock on product packages. Analysis of the data

shows that 3.03 million units of twelve non-MSL labeled products were sold during the sixty-five week period in the London market area, while 3.3 million units of twelve MSL-labeled products were sold during the same period.

1.4 Model Specification and Estimation

The hedonic model specifies the price of a product as a function of the attributes that characterize the product. The model can be written in its general form as:

$$P_{it} = f(s_1, \dots, s_n)' \tag{1.1}$$

where P_{it} is the price of good i at time t , and $S = (s_1, \dots, s_n)$ is a vector of attributes that determines the price of the good. As such, the function allows a test of the value of each attribute, given that all other attributes are held constant. Each attribute j can be measured on a continuous scale or by a dummy variable depending on its type.

Linear and log-linear functional forms are frequently found in the literature for such models. For example, McConnell and Strand [61], Carroll, Anderson, and Martinez-Garmendia [62] and Smith, Huang and Lin [47] use linear functional form for hedonic models of individual products pricing. Smith, Huang and Lin [47] cite ease of interpretation as one reason to use the linear model in their estimation of the premium for organic milk, as the parameter estimates on the organic attribute variable can be interpreted as the premium in currency units. Similarly, a log-linear functional form can also be easily interpreted, where the parameter estimates are evaluated in percentages. In this application, a multiple of products of different base values are included in the analysis. A Box-Cox test showed no functional form was clearly preferred, thus, given that a percentage over the base value is more meaningful a

log-linear functional form was specified :

$$\ln P_{it} = \alpha + \sum_{j=1}^k \beta_j \text{brand}_{jit} + \sum \phi_l \text{form}_{lit} + \sum \gamma_n \text{type}_{nit} + b \ln \text{Size}_{it} + c \text{MSC}_{it} + e_{it} \quad (1.2)$$

where j indexes the brand attributes (Youngs, Birdseye, or private label), l indexes process form (breaded, battered, or natural/smoked), n indexes product type (fish fingers, fillets, or kids fish) and e_{it} is a random error. In this analysis, the attributes are all expressed as dummy variables (Table 1.1), except package size. Package size is specified as a continuous variable, in logarithmic form. Dummy variable coding was used instead of alternative coding, such as effects coding, as it is easily interpretable given the large number of attributes included in the model specification². This follows established methodology of previous literature using hedonic methods to estimate the marginal value of attributes of seafood products, including McConnell and Strand [61], Carroll, Anderson, and Martinez-Garmendia [62], and Roheim, Gardiner and Asche [65], as well as hedonic analyses of organic produce [46], organic milk [47], organic tomatoes and apples [48], and ecolabeled apparel [66].

By including a constant term, the parameters are interpreted as the percent deviations from a basic product with a given set of attributes. In each dimension one can investigate whether the different attributes have different marginal values by testing whether the associated parameters are zero. Own-label, kids fish, natural smoked, and non-MSC labeled serve as the base attributes for the model. Models which included interactive effects between the MSC label and other attributes were tested, however none were statistically significant.

Since scanner data contain observations on multiple products of differing average values, the variances of the error terms are likely to differ across products. Whites

²Dummy variable coding was primarily chosen to ensure that the model results would be directly comparable with related work that has estimated price premiums for the MSC ecolabel on products derived from other fish species [63, 64].

[67] test rejected the hypothesis of homoskedasticity at the 1% significance level. The model, which was run using *STATA*TM, corrected for heteroskedasticity with a heteroskedasticity-consistent covariance matrix estimator [68]. Following Davidson and MacKinnon [69], the HC3 covariance matrix estimator was used. The data were also tested for the presence of multicollinearity, although no significant effect on model results was detected.

1.5 Results

Table 1.2 reports the coefficient estimates and goodness-of-fit of the model. Overall, the estimated equation is highly significant with a p-value < 0.0001 and t-statistics indicate each coefficient is significant at a minimum of the 5% level. These results show the heterogeneity of the seafood market. The individual coefficients indicate that, on average, Birds Eye branded frozen processed pollock products are 56% more expensive than private label products, while Youngs branded products are 35.5% more expensive than private label products. A more precise percentage premium results from subtracting 1 from the exponential of the parameter, as that represents the proportionate difference, holding all other attributes constant [70]. As a result, the Birds Eye brand products are 76% more expensive than private label products, while Youngs brand products are 43% more expensive. To determine if branding contributed to the fit of the equation as an attribute category, an F-test was performed and reported in Table 1.3. Significant at the 1% level, this test rejects the hypothesis that a product price is independent of brand, indicating that brand is an attribute that contributes to the segmentation of the seafood market.

Breaded pollock products are 11% less expensive than natural smoked pollock, while battered is 24% less expensive and both are individually statistically significant. While typically considered value-added products, breaching and battering may be adding value to a product which is of lower value from an initial state, perhaps because

of lower quality. In other words, if the product were of sufficiently high quality, one might expect that the fish would be marketed as the higher-valued product, natural. Thus, so-called value-added from breaching and battering actually may be a process form that masks some of the quality control issues generated in the supply chain. Again the F-test result shown in Table 1.3 indicates that process form, as an attribute sub-group, significantly explains changes in product price which follows intuitively from the previous discussion.

Filletts are 49% more expensive than kids fish, while fish fingers are 6% less expensive than kids fish, statistically significant both individually and as an attribute group. There is a positive and significant relationship between price and portion size.

The focus of this analysis is whether or not there is a price premium for MSC-labeled products. Thus, the premium is estimated to be 14.2% on these MSC-labeled processed frozen processed pollock products relative to non-MSC labeled frozen processed pollock products after fully accounting for the other product attributes such as brand, product form, package size and process form.

While it is useful to put this estimated premium into context, there are limitations in our ability to do so. First, as mentioned previously, there are no existing studies of actual premiums paid for ecolabeled seafood in the UK market or any other market, or for any other seafood products.³ Secondly, previous analyses of consumer preferences for ecolabeled seafood have not generally estimated willingness to pay (WTP), but rather evaluated factors which influenced the probability of hypothetical purchase of ecolabeled seafood products, including Wessells, Johnston and Donath [4], Johnston et al. [5], Johnston and Roheim [6], Jaffry et al. [3], Brcard et al. [7], and Salladar et

³Since this research has been published, there have been two subsequent studies that use Hedonic analysis to estimate the price premium for the MSC ecolabel. The first [63] finds a similar estimate of 10% premium for the MSC label on chilled, pre-packed Cod and Haddock and the second [64] finds a 13.1%. Both studies utilize data from several supermarket chains in the UK. The results from this study support their findings.

al. [37]. Only in Johnston et al. [5], in an international comparison of roughly 2,000 consumers in both the U.S. and Norway, a within-sample prediction was performed that showed 80% of U.S. consumers would be willing to pay an average 24% premium for ecolabeled salmon, cod and shrimp, while 54% of Norwegian consumers would be willing to pay the same price premium. These estimates are higher than the actual premium estimated as paid in the UK market, and may be a result of the hypothetical nature of the survey-based study in Johnston et al. [5] as well as differences in geographical markets.

Other interesting comparisons may be to look at alternative forms of product differentiation. Focusing on analyses that use scanner data and hedonic methodology to statistically estimate actual price premiums, we investigate existing literature for organic, fair trade, and branding attributes. Among these, Galarraga and Markandya [45] find an 11% premium in the UK market for fair trade coffee over regular coffee. Roheim, Gardiner and Asche [65] determine the value of branding finding that national brands across many seafood commodities in the UK have a 10% premium over private labels. Lin, Smith and Huang [46] and Smith, Huang and Lin [47] show that organic labeling in the U.S. yields price premiums between 15% and 60%, depending upon food product and geographical market within the U.S. This implies that seafood ecolabels may be valued slightly higher than fair trade coffee in the UK. However, the premium is on the lower end with respect to what has been reported for organic products in the USA. Such a difference in premiums may not be surprising as fair trade and environmental sustainability may yield only warm glow effects which may be less welfare improving in terms of consumer utility than the combination of environmental sustainability and perceived health benefits potentially derived from consumption of organic products [71, 72].

1.6 Conclusions

Success of ecolabeling programs in fisheries depends upon: a) a sufficient number of well-managed fisheries becoming and remaining certified, thus placing a critical mass of certified product into the supply chain; b) creating the incentives to reform poorly managed fisheries such that they become well-managed fisheries. To create that success, market benefits are necessary for ecolabeling programs to influence production and management practices in any industry. Price premiums are a direct means by which to offset costs incurred from sustainable fishing practices certified under fisheries ecolabeling programs, and are more directly measureable than other market benefits such as improved market access or expanded market share. To date, all evidence of the effect of ecolabels for seafood has been obtained using survey data [4, 5, 3, 6, 37]. Anecdotal evidence indicates the shift of European processors such as Unilever, Youngs Bluecrest and Frosta from sourcing Russian pollock toward U.S. pollock due to sustainability certification [54]. However, doubts have frequently been expressed that price premiums actually exist [43, 42]. In relation to the MSC, Washington [43] stated that the price premium is a myth and the OECD [42] stated that no evidence exists which documents effectiveness of ecolabeling schemes in creating market incentives for better fishing practices. Data limitations and complexities of the market often make it difficult to quantify market benefits [73, 35].

This paper provides statistically significant ex post evidence of market benefits of fisheries certification at the retail level: payment of price premiums for ecolabeled seafood, possibly the first analysis to do so. A limitation of the research lies in its scope it focuses only on the London metropolitan area, and on frozen processed pollock. There are several remaining questions. For example, is the premium sufficient to cover costs of a sustainable fishery and certification? This remains an open question, as the UK market represents only a small portion of the market into which Alaska

pollock is sold. Further research is needed to determine the size of the premium in the other markets for pollock (the rest of Europe, North America and Japan). Furthermore, yet to be determined is whether the premium transmits from the retail level to the production level to compensate those who are engaged in fishing activities. Indeed, it is not clear that there is adequate information for comparison on the actual cost of transition from a fishery that does not meet the criteria of sustainability to one that does, in addition to the costs of certification and maintenance of certification. In addition, over 100 other certified fisheries and many other markets are yet to be investigated. The most compelling evidence that benefits outweigh the costs comes from the behavior of the fishing industry itself: the number of fisheries in assessment and becoming certified continues to grow. Thus, this research can only be the beginning of the analysis of the market benefits of MSC certification. Yet it is a reasonable beginning, and does show that a price premium is being obtained at the retail level for certified sustainable pollock over non-MSC labeled pollock which does not carry the MSC label, holding other product attributes constant.

Table 1.1. Variables and Their Descriptive Statistics

Variable	Description	Mean	Std Dev
Price	Average unit price in pounds sterling	2.02	0.94
Ln P	Log of Price	0.58	0.52
Size	Package Size in grams	436	217
Ln Size	Log of Size	5.98	0.43
Birds Eye	Dummy Variable for Birds Eye brand, 1 if present, 0 if not	0.46	-
Youngs	Dummy Variable for Youngs brand, 1 if present, 0 if not	0.39	-
OwnLabel	Dummy Variable for Own Label brands, 1 if present, 0 if not	0.15	-
MSC	Dummy Variable to indicate presence of MSC logo, 1 if present, 0 if not	0.51	-
Breaded	Dummy Variable to indicate if product was coated in breading, 1 if so, 0 if not	0.49	-
Battered	Dummy Variable to indicate if product was coated in batter, 1 if so, 0 if not	0.39	-
Natural Smoked	Dummy Variable to indicate if product was natural smoked, 1 if so, 0 if not	0.01	-
Fillet	Dummy Variable to indicate if product was in fillet form, 1 if so, 0 if not	0.48	-
Finger	Dummy variable to indicate if product was in fish finger form, 1 if so, 0 if not	0.4	-
Kidsfish	Dummy variable to indicate if product was in kidsfish form, 1 if so, 0 if not	0.11	-

Standard deviations not reported for dummy variables

Table 1.2. Parameter Estimates

Variable	Coefficient	Std. Err	t-ratio
Intercept	-4.79***	0.13	-36.93
MSC	0.133***	0.023	5.77
lnSize	0.822***	0.02	42.09
Youngs	0.355***	0.044	8.06
BirdsEye	0.564***	0.033	16.9
Breaded	-0.116***	0.044	-2.62
Battered	-0.273***	0.031	-8.77
Fillet	0.398***	0.031	12.74
Finger	-0.067**	0.029	-2.32
R^2		0.649	
No. of observations		1137	
Pr>F		p<0.0001	

Standard errors in parentheses; *** indicates $p < 0.01$, ** indicates $p < 0.05$. Standard Errors are HC3 Robust Standard Errors.

Table 1.3. Comparison of hypotheses of attribute category inclusion

Null Hypothesis	F(2, 1128)	Prob > F
Brand: 1 = 2 = 0	165.50***	0
Process Form: 1 = 2 = 0	78.29***	0
Product Type: 1 = 2 = 0	149.36***	0

*** indicates $p < 0.01$.

Factors that influence response reliability in survey data: Addressing challenges of employing incentive-compatible mechanisms outside the lab

Prepared for submission to the Journal of Choice Modelling

Julie Insignares Santos⁴

⁴PhD Candidate, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881. Email: jinsignares@my.uri.edu

MANUSCRIPT 2

Factors that influence response reliability in survey data: Addressing challenges of employing incentive-compatible mechanisms outside the lab

2.1 Introduction

Many ecosystem services enjoyed by local communities from farmland are public goods and thus are likely under-provided. For instance, agricultural land may provide wildlife habitat, scenic views, preservation of bucolic character, and groundwater aquifer recharge, all of which undoubtedly have some value to the local community. To enhance these ecosystem services, there is a growing interest in market approaches in which those who value the services pay those who can provide them at the least cost [74]. However, constructing markets for public goods is complicated by free-ridership. Because it is often prohibitively difficult to preclude non-payers from benefiting from the good, buyers do not have the incentive to participate in the market or pay their full values given that they cannot be excluded from consumption.

Historically, field experiments involving public goods rely upon voluntary contribution mechanisms to procure funds. However, decades of experimental economics outcomes have proven voluntary donations to be poor at achieving aggregate demand revelation. Given this flaw, researchers have developed alternatives to the voluntary contribution elicitation mechanism that have led to significant improvements in aggregate contribution. For instance, several authors [8, 9, 10] have shown that establishing a provision point with some form of rebate rule and money back guarantee if the provision point cannot be met can significantly reduce free-riding. In fact, the pivotal mechanism, which promises to collect payment only if, given the contributions of others, the participant's offer is required to ensure provision of the good, has the advantage of incentive compatibility in certain circumstances. That is, theoretically,

revealing one’s full value for the good truthfully is a dominant strategy.

The question arises: how will the mechanisms that help alleviate free-riding transfer from the laboratory to the field? Because they are often complex, unfamiliar in use, and require unanimous action or multiple rounds of interactions, several authors have pointed out that it may be difficult to extend the benefits of the elicitation mechanisms to the field [8, 11]. In practice, there are several pertinent differences between laboratory and field environments that would suggest a need for careful consideration of public goods mechanisms in the field. First, participants in laboratory experiments often commit to a set duration of time, during which the researcher can ensure that directions have been read carefully. There are often several practice rounds administered to participants during which questions, if they arise, can be addressed. In many field experiments, lengthy instructions are provided but there is no guarantee that respondents read or comprehend them in their entirety.

We examine this issue by utilizing a novel data set involving one of the first efforts to bring several of the most promising public goods elicitation mechanisms to the field. In 2006, the researchers at the University of Rhode Island, in collaboration with EcoAsset Markets, Inc., implemented a choice experiment describing a hypothetical market for wildlife preservation in Jamestown, Rhode Island. The “good” offered for exchange was a contract between the residents and the farmers within the community that would change hay harvesting practices to protect grassland nesting bird habitat. Jamestown is a small community located off the coast of Rhode Island on Conanicut Island in Narragansett Bay. There are nine farms on the island, most of which produce grass-fed beef. There has in the past been evidence that the Jamestown community places a high valuation on its farms [75]. In addition, its residents tend to have a keen sense of attachment to their community and a history of supporting conservation of low-impact land use[76, 77]. These characteristics made Jamestown an ideal venue

for which to test a local market for ecosystem services. The choice experiment was conducted to measure both the demand for the protection of grassland nesting habitat and how this demand varies across several public goods auction mechanisms.

In the absence of direct feedback about how individuals respond to the mechanisms they are administered, we evaluate choice consistency as a measure of the complexity of the choice task. Researchers have used econometric models that explicitly incorporate heteroskedasticity in error variance (scale heterogeneity models) to measure choice consistency in the presence of choice situation complexity, learning and fatigue effects [23, 78, 22, 20, 21]. In addition, scale heterogeneity models have been used to compare survey elicitation methods. Open-ended, discrete choice, or payment card formats have been examined as well as differences between contingent valuation and choice experiment surveys[17, 79, 80]. Rather than vary the survey format or the number of alternatives, choices, or attributes, we focus on the elicitation mechanisms administered and hypothesize that more complex mechanisms lead to greater choice inconsistency. This study will test whether more complex or unfamiliar mechanisms of payment complicate the decision process sufficiently to cancel out the benefits from incentive-compatibility. By comparing the variance in response across individuals, we identify quixotic behaviors and test for linkages within and across mechanism treatments. In combination with individual-specific WTP measures, we are able to draw valuable conclusions about whether the mechanisms are performing as expected. In this manner, this work contributes to the burgeoning literature on market design for ecosystem services. Additionally, we supplement the growing literature regarding the nature of scale heterogeneity[81].

Using the data from the choice experiment, we measure the randomness in subjects' responses by estimating individual-specific scale coefficients from a mixed logit model (MIXL) in willingness-to-pay space using hierarchical Bayes procedures. De-

mographic determinants of the scale parameter values are explored as well as choice task features and strategic behavior such as yea-saying and lexicographic response. Examining choice consistency is particularly relevant to the data generated from choice experiments involving public goods elicitation mechanisms because individuals who participate in the choice tasks must undertake two somewhat complicated thought processes which would require significant cognitive effort. First, the choice experiment invites participants to consider the provision of environmental amenities as salable goods with a variety of attributes that would be individually considered for valuation. In addition, respondents would then consider how their response would be processed given the elicitation mechanism administered.

In addition to the main goal of testing how subjects' scale parameter estimates are affected by the complexity of elicitation mechanisms, another novel contribution of this paper is that we specify a WTP-space version of the random-scale multinomial logit model and test it against a full model, in which both scale and attribute parameters are allowed to vary. The scale heterogeneity model derived by Louviere and Eagle [81] assumes that all heterogeneity in preferences can be adequately explained by variations in scale alone. However, Adamowicz et al. [23] point out that accounting for both sources of heterogeneity is particularly important in welfare estimation, new product design, and segmentation in marketing. We add to the discussion by supporting the finding that scale heterogeneity is particularly important when attributes are highly correlated.

The layout of the paper is as follows. The next section describes the Jamestown choice experiment in detail along with the elicitation mechanisms that were administered. The three main hypotheses regarding the elicitation mechanisms are outlined. The third section describes the WTP space approach and presents the model specifications. Section 2.4 describes the mechanics of the estimation with emphasis on

the steps in the MCMC procedure. Section 2.5 presents results of the HB estimation along with the second stage scale regression. Section 2.6 concludes the paper and suggests further research.

2.2 The Experiment

In order to facilitate the design of a local market for ecosystem services, it was important to choose an ecosystem service that could be easily quantified, implemented on a sufficiently short time line, and be relatively inexpensive. Agricultural land provides support for many important and valued wood-edge species. It was determined that wildlife habitat preservation might fit the requirements for this type of experiment. Specifically, the black and yellow Bobolink (*Dolichonyx oryivorus*) utilize hay fields in Jamestown, RI as nesting habitat during a five week interval spanning the months of May and June. Hay harvesting and grazing activities during this period prove devastating to cohort success [82]. As a vulnerable species that would benefit from conservation efforts, the Bobolink cohort could be salvaged by a modest shift in harvesting practices.

Each choice situation described in the experiment contained hypothetical contracts under which the farmers would agree not mow or graze on the contracted acreage until after the fledgling Bobolinks have had time to mature. The contracted acreage could be seen from the road or not and potentially supplemented with an invitation to an expert-led birdwalk and additional fallow acreage to be restored to active farming. Consultations with biologists indicated that delaying harvest until after July 4th would achieve this goal.

The survey was administered between October and December 2006 by mail to all valid addresses in Jamestown, RI, a total of 2893 households. After deducting undeliverable addresses, the response rate was 38.2%. There were 791 respondents in the final analysis. There were 10% more female respondents than male. The mean

age of respondents was 57 years of age. Over 73 percent of respondents indicated that they did not have children under the age of 18 in their household. The median level of education of respondents was some college with a median level of income between \$100,000 and \$199,000 [77].

The survey was comprised of five sections. The first section described the ecosystem service in question. The survey described how Bobolinks utilize hayfields in Jamestown and how current harvesting practices impact breeding and rearing. This section also provided information about other important environmental services that hayfields provide. The second section included the choice experiment itself, with six questions regarding two hypothetical farm-wildlife contracts and a no-buy option. Before being asked to choose between contracts, respondents were given information about how their elicitation mechanism worked. Respondents were given the option to choose one of the two contracts, neither, or both. The sixth question was unique in that it consisted of only one contract and respondents were asked to indicate whether they would purchase the contract or not. The remaining sections solicited participants for their opinions regarding farmland amenities and rural community character and collected demographic information.

The choice experiment included six attributes of the farm-wildlife contracts (Table 2.1). The first attribute was the ‘Acres of managed hayfields’. This attribute also incorporated the expected number of bobolink fledglings. In estimation, this attribute is called ‘High Bobolink’ and has two levels: high and low and was mildly correlated with the number of acres of contracted hayfields. The second defining characteristic of a contract was ‘acres of restored fields’. Preliminary discussions with the farmers residing in Jamestown suggested that there may be opportunity to restore fallow land to active hay production. The ‘View’ and ‘Tour’ attributes were both binary attributes and represented whether the parcel was viewable from the road and whether

the contract came with an invitation to an expert-led birdwalk. This last attribute was unique in that it was the only attribute that can be viewed as a purely private good characteristic. The final attribute was the contract's cost: an eight level attribute ranging from \$10 to \$105. An example of a choice experiment scenario is included in Appendix A. The scenarios were constructed from a $4^2 \times 2^3 \times 8 \times 2$ orthogonal main effects design and blocked into groups depending on the treatment.

The Elicitation Mechanisms

There were two main groups of surveys administered. The main groups differed in whether a reference mechanism was applied or not. The reference mechanism was a hypothetical referendum for a tax increase. Individuals who were assigned to this group were presented with two choice situations in which the mechanism of revenue collection was an increase in taxes in the amount of the cost of the contract. The intention of providing the reference mechanism was to test whether respondents who were administered a familiar base mechanism would then be more responsive to a less familiar mechanism. We test the significance of providing a familiar reference mechanism on the behavior of the respondent.

Hypothesis 1 *Providing a familiar reference mechanism highlights the advantages/disadvantages of the alternative mechanisms and so individuals will be more responsive*

Each of the two main groups was further divided into several subgroups that differed in the elicitation mechanisms that were administered. We tested four elicitation mechanisms. All four included a provision point and money-back guarantee: if not enough revenue was collected to compensate the farmers, all of the money would be returned to the respondents. The value of the provision point was not revealed to the participants. The first elicitation mechanism, the provision point with money back guarantee (PP/MBG), collects all offers unless there are not enough to reach

the provision point. This mechanism was only included in the treatment that was administered a reference mechanism. The second mechanism, the provision point with proportional rebate (PR), builds upon the PP/MBG by offering a proportional rebate of excess contributions. While the PP/MBG and PR mechanisms have shown evidence of alleviating some free-riding behavior [9, 8, 10], they are not in theory incentive-compatible. The pivotal mechanism (PM) was tested as the third mechanism because it has been shown to be incentive-compatible in controlled experiments. The pivotal mechanism promises to collect from the participant only if her contribution makes the difference between meeting the provision point or not. In this manner, the pivotal mechanism induces the respondent to consider revealing her true value for the good. In the Jamestown choice experiment, PM came with a provision point and money back guarantee as well. However, Milgrom [11] points out that incentive-compatible mechanisms tend to be more complex and may be difficult to implement outside of the lab. Our second hypothesis is that, if individuals are not responding as predicted to the pivotal mechanism, this aberration can be explained by a smaller scale parameter (larger error variance), thereby indicating that the complexity of the choice situation has overtaken the advantages of the mechanism.

Hypothesis 2 *Relative to the other public goods elicitation mechanisms tested in this experiment, the performance of the pivotal mechanism outside the lab is hindered by its complexity.*

The fourth mechanism tested in this experiment was the Uniform Price Auction (UPA). Under UPA, the respondent is informed that after all bids are collected a uniform price will be determined such that all bidders pay the same price for the good, provided her bid was above the determined price. This mechanism has a fairness aspect to it that the others lack. Fairness has been a prime motivator in many cases in the lab[83]. In addition, paying the same price for a good is a common

occurrence in markets for private goods and hence the expectation that everyone pays the same price for the good may be appealing.

Hypothesis 3 *The Uniform Price Auction is likely to succeed in achieving higher demand revelation and more consistent choices than the PP/MBG and PR mechanisms because the aspect of fairness is appealing to respondents and resembles a private market setting.*

In summary, there were 256 pairs of farm-wildlife contracts and 32 single contracts. These were divided among four main groups based on whether the reference mechanism was applied and then further subdivided into groups based on auction mechanism. The mechanism descriptions that were administered in the survey can be found in Appendix A.

2.3 The WTP-space approach

Traditionally, choice experiments have been estimated via classical mixed logit applied to random utility theoretic models of preferences.¹ Suppose an individual $i = 1 \dots I$ faces a choice instance defined by c alternatives, in this case three and the “no buy” alternative. Utility is assumed quasilinear and separable in price. The utility to individual i of choice c among alternatives has both a random and deterministic component and takes the following form:

$$U_{ic} = -\alpha p_{ic} + \beta' x_{ic} + \varepsilon_{ic} \forall c \quad (2.1)$$

¹ The basis of random utility theory originates with Lancaster [84] and Rosen [85] and assumes that utility for a good can be derived from the characteristics of the good. Mixed Logit models account explicitly for individual-specific heterogeneity in tastes and avoid the limitations of IIA restrictions of the simple multinomial logit model. In the transportation and environmental economics literature alone, there have been a preponderance of studies incorporating the mixed logit model. Carlsson et al. [86] utilize the mixed logit model to estimate preferences for wetland attributes in southern Sweden. Bhat [87] estimate the mixed logit model to analyze choice of inter-city travel. Sillano and Ortuzar [88] use random-parameter mixed logit to estimate willingness to pay to avoid poor air quality alerts. Train and Sonnier [89] model consumers’ choice among gas, electric, and hybrid cars.

where p_{ic} is the price of the contract faced by individual i in choice situation c and x_{ic} is a vector of attributes of the contract including: number of acres protected, number of acres restored to active farmland, whether the plot has a view from the road, and whether the contract provides buyers with an opportunity to attend an expert-led birdwalk (we exclude the number of bobolink territories from the analysis because it was not shown to add significantly to the estimation). In the traditional MNL specification, α and β are homogeneous in the population and represent the marginal utility of income and marginal utility of the attributes of the contract, respectively. ε_{ic} captures the unobserved factors that influence the utility of person i for alternative c . In order to obtain choice probabilities that are in the set $[0,1]$, we assume that ε_{ic} is Gumbel distributed with $\text{Var}(\varepsilon_{ic}) = \sigma^2(\pi^2/6)$. σ^2 represents the variance of the unobserved factors that influence utility. The standard deviation of σ^2 is termed the scale of utility. This term is not a component of utility itself, but represents the standard deviation of the random portion of utility. If we define the original error term ε^* , then $\text{Var}(\frac{\varepsilon^*}{\sigma}) = (\frac{1}{\sigma^2})(\sigma^2)(\frac{\pi^2}{6})$. The choice probabilities then become

$$P_{ic} = \frac{e^{(-\frac{\alpha}{\sigma}p_{ic} + \frac{\beta'}{\sigma}x_{ic})}}{\sum_j e^{(-\frac{\alpha}{\sigma}p_{ij} + \frac{\beta'}{\sigma}x_{ij})}} \quad (2.2)$$

σ is not separately identifiable from marginal utility parameters and so in general it is fixed to 1 and parameters $\alpha^* = \frac{\alpha}{\sigma}$ and $\beta^* = \frac{\beta}{\sigma}$ are estimated. This reparameterization is of little concern if the aim of the estimation is ultimately to derive measures of marginal rates of substitution such as willingness to pay measures as these values generally require dividing one parameter (the marginal utility of an attribute) by another (the marginal utility of income or the cost parameter). In these cases, the scale factor cancels out.

However, the implicit scale parameter becomes a problem in two cases. First, scale of utility is a concern when comparing coefficients between groups. In general, larger scale implies smaller coefficients overall, even if the underlying preference parameters

are the same across groups. A second issue, and one that is central to this paper, is that once heterogeneity is incorporated into the model, then discerning variation in tastes from variation in scale is not possible. That is, once we move from a model that does not model heterogeneity to one that does, then assuming that the standard deviation of error variance is the same for all respondents is a strong assumption and one that, in most cases is not supported.

To address the shortcomings of the preference approach, we follow the approach popularized by Train and Weeks [24] and others [90, 91, 92, 93, 94] and estimate the WTP-space specification.² Instead of choosing the alternative that maximizes utility, the consumer chooses the alternative that maximizes consumer surplus, which is the difference between her reservation price for that alternative and its price. Therefore she chooses alternative c if

$$R_c - p_c \geq R_j - p_j \forall j \quad (2.3)$$

where R_c is the respondent's maximum willingness to pay for alternative c and p_c is c 's price, and this difference is a measure of consumer surplus. In this case, the logit choice probability becomes

$$P_{ic} = \frac{\exp\left[\frac{\beta'x_{ic} - p_{ic}}{\sigma}\right]}{\sum_j \exp\left[\frac{\beta'x_{ij} - p_{ij}}{\sigma}\right]} \quad (2.4)$$

In this specification, the β s are interpreted as direct measures of marginal willingness to pay and the scale parameter is explicitly estimated.³

² Jedidi et al. [95] outline a proof of equivalence between surplus and utility maximization.

³ Fiebig et al. [96] and Sonnier, Ainslie and Otter [97] point out that the two models differ only in the restrictions placed on parameters. That is, by fixing the scale parameter of the WTP-space model to 1 and estimating a coefficient on price, the preference space model is obtained. Similarly, by normalizing the cost coefficient and estimating the scale parameter in the preference space model, one obtains estimates in surplus space. It is worth pointing out that fixing the cost coefficient in surplus space is equivalent to restricting the marginal utility of income to be constant only if the researcher restricts the WTP coefficients to be homogeneous in the population. In a random parameter model, however, variation in the marginal utility of income is incorporated in the variation of the WTP coefficients, just as random parameters in preference space incorporate both random taste heterogeneity and random scale heterogeneity. In the model in WTP-space, it is not possible to separate heterogeneity in the marginal utility of income from heterogeneity in the

Using the WTP-space approach, we examine the sources of heterogeneity in choice consistency. Individuals can exhibit response heterogeneity in choice experiments in several different ways (Table 2.2). First, the alternative-specific constants (ASCs) represent general differences in preferences across alternatives. If these are allowed to vary, then we allow each individual to have different patterns of residual taste heterogeneity across alternatives. Second, as discussed previously, the scale of the error term, or the standard deviation, is a representation of heterogeneity in response across choices for a particular respondent. Allowing the scale term to vary across individuals accommodates heterogeneity in the consistency of choices by respondent. Finally, the regressors on the attribute variables can be held fixed or allowed to vary by respondent. This type of heterogeneity is accounted for in the standard random-parameter logit model, and is generally termed taste heterogeneity.

Among these sources of heterogeneity, we posit that the scale parameter largely represents the heterogeneity in response in our application. Heterogeneity in scale parameter can come from choice set design, respondent characteristics as well as product complexity[81]. Train and Weeks [24] posit that scale heterogeneity may arise because of purely idiosyncratic behavior on the respondent’s part, or through differences in the variance of unobserved factors over choice situations.

Several studies support the view that the scale parameter captures much of the response heterogeneity. Louviere and Eagle [81] posit that much of the heterogeneity encountered in choice models can be accounted for simply by modeling scale heterogeneity. In this sense, the scale parameter varies but attribute coefficients are fixed. Thus, respondents are assumed to have the same preferences but these preferences are shifted up or down based on the individual-specific scale of the error term. Fiebig et al. [96] analyze the contribution to improvements to loglikelihood gained from

marginal willingness to pay for an attribute. In the model in preference space, it is not possible to separate heterogeneity in scale from heterogeneity in preferences.

sequentially adding sources of heterogeneity. They find that the biggest improvement in log-likelihood from incorporating scale heterogeneity can be found in the data sets that involved rather complex goods, such as medical decisions and cell phones as opposed to more mundane consumer goods such as pizza delivery and vacation destination.

On the other hand, Adamowicz et al. [23] point out that accounting for both preference and scale heterogeneity is particularly important in welfare estimation, new product design, and segmentation in marketing. Given these contrasting views, we specify a WTP-space version of the scale heterogeneity model and test it against a full model, in which both scale and attribute parameters are allowed to vary. To our knowledge, this study is the first to conduct this test.

Heterogeneous response in our application may arise from three sources. First, as a new product for this community, with no comparable substitutes, differences in the variance of the error term across individuals may indicate that some individuals have a clear sense of their values for the good from the start while others are uncertain. Second, Fiebig et al. [96] find that scale heterogeneity matters most in cases where there are “extreme” respondents who may make decisions that are not consistent with random utility maximization. These individuals exhibit behaviors that include lexicography, protest votes, and yea-saying. By inspecting the data, we were able to find evidence of all three response strategies. The third source of scale heterogeneity involves possible reactions to the mechanism treatments themselves. Systematic differences in scale across treatments indicate differences in the average level of idiosyncratic behavior measured by the model. In this light, differences in scale indicate the extremity of reaction to changing the rules of the game regarding payment.

Model Specification

The analysis is based on four contract characteristics (Acres, Restore, View, and Tour), cost, and two alternative-specific constants (ASCs). The fifth contract attribute, High Bobolink, was mildly correlated with the Acres attribute and was excluded from estimation because it was not found to significantly contribute to the explanatory power of any model specification. The two alternative-specific constants represented the No-Buy option and the Both option.

In an effort to isolate taste and scale heterogeneity, three models were estimated. The first model assumes that behavior can be sufficiently captured by taste heterogeneity alone. The second tests whether behavior is best described by scale heterogeneity and the third model incorporates both taste and scale heterogeneity.

There are a few considerations with regard to model specification. First, Fiebig and colleagues [96] point out that scaling the alternative-specific constants leads to complications in estimation if there is a significant fraction of the population that always chooses the same alternative. In our application, scaling the ASCs exacerbated autocorrelation in the MCMC chains. In fact, a substantial proportion of the sample (n=109/791 respondents) chose the Both alternative for all questions. Therefore, the ASCs were left un-scaled so as to facilitate model convergence.

The second consideration is generally cited as one reason why the cost parameter in Utility-space specifications is modeled as a fixed parameter in the population. It is often the case that a fully random specification is empirically intractable. We found this to be the case for our data as well and, hence, at least one parameter is specified as a fixed parameter in all three estimations.

The final taste heterogeneity model is formulated as follows:

$$V_{njt} = \alpha_n ASC_{BOTH} H_{njt} + \gamma_n ASC_{NO} O_{njt} + \sigma \beta_n x_{njt} - \sigma cost_{njt} + \varepsilon_{njt} \quad (2.5)$$

where V_{njt} denotes the value function of individual n for alternative j in choice situation t . x_{njt} is the vector of contract attributes and β_n are the individual-specific marginal WTP estimates. Upon inspection of the MCMC chains, there was strong evidence that the hyperparameters are highly correlated. Survey responses support this finding: a large majority of respondents exhibited homogeneous preferences with regard to the contract attributes. Based on this information, we aimed to test the hypothesis that differences in behavior for this sample can best be described by scale heterogeneity alone. The value function for this second model is

$$V_{njt} = \alpha_n ASCBOTH_{njt} + \gamma_n ASCNO_{njt} + \sigma_n \beta x_{njt} - \sigma_n cost_{njt} + \varepsilon_{njt} \quad (2.6)$$

Unfortunately, because the scale parameter is inextricably linked to the ASCs, it was not possible to restrict the ASCs to be fixed in the population while allowing for scale heterogeneity. This is best exemplified by examining a choice of Both contracts in a choice situation. By choosing both contracts, an individual is effectively indicating that she is not willing to make any trade-offs among attributes. Therefore, the influence of the attribute characteristics themselves in the value function decreases. This is achieved by a lower value for the scale parameter.

The third model incorporates both taste and scale heterogeneity. However, because specifying a fully random model was empirically infeasible, there was a need to restrict at least one parameter to be fixed in the population. There was not sufficient response heterogeneity to accommodate random parameters for both the View and Tour attributes. Therefore, these were specified fixed parameters.

$$V_{njt} = \alpha_n ASCBOTH_{njt} + \gamma_n ASCNO_{njt} + \sigma_n \beta_{n,Acres} Acres_{njt} + \sigma_n \beta_{n,Restore} Restore_{njt} + \sigma_n \beta_{Tour} Tour_{njt} + \sigma_n \beta_{View} View_{njt} - \sigma_n cost_{njt} + \varepsilon_{njt} \quad (2.7)$$

We utilize the three models in order to determine how much of the heterogeneity in response can be accounted for by scale alone and how much incorporating taste

heterogeneity adds to the explanatory power of the model.

2.4 The Model and Estimation

We estimate the model using hierarchical Bayes (HB) estimation procedure was chosen for several reasons. First, our research draws heavily on outcomes at the level of the individual. HB incorporates these calculations into the overall estimation quite efficiently. In contrast, classical methods treat individual-specific estimates as an afterthought. Fiebig et al. [96] point out that an advantage of HB is that specifying diffuse normal priors for the individual-level parameters allows for highly flexible posterior distributions. In addition, we expect the parameters of the attributes of the alternatives to be highly correlated based on supplementary sections of the survey which indicated that most individuals placed positive values on the attributes. Therefore, estimating a model that restricts the covariance matrix to be diagonal would be a serious misspecification. Bayesian estimation incorporates full covariance matrices far better than classical estimation[98]. Das et al. [91] find that accounting for correlated coefficients in a model estimated by classical methods slows down calculation considerably. Finally, we recognized the limitation in data per person: the number of attributes under consideration was just short of the number of choice situations faced by the individual. Allenby and Rossi [99] point out that hierarchical Bayes methods are particularly well-suited for data that consists of many individuals with relatively sparse information per unit of analysis, or “short” panels.

The Mechanics of HB Estimation

Bayesian estimation begins with the assumption that we can combine our expectations with our observations to update our beliefs about the world. We represent our expectations independent of our observations by prior distributions on the parameters to be measured. What we observe is represented with a likelihood function. These two combine to produce the posterior distribution of the parameters which is

a weighted average of the two. Formally,

$$Posterior \propto Likelihood * Prior \quad (2.8)$$

For the HB estimation of the WTP-space model with fixed and random parameters, we can represent this relationship as:

$$K(b, W, \beta_n \forall n | Y) \propto \prod_n L(y_n | \beta_n) \psi(\beta_n | b, W) k(b, W) \quad (2.9)$$

where b is the vector of means of the population-level parameters, W is the variance-covariance matrix, β_n represents the individual-specific estimates of b , Y are the observed choices, $\psi(\beta_n | b, W)$ and $k(b, W)$ are the prior densities on the parameters, and $L(y_n | \beta_n)$ is the likelihood function for each person. The likelihood function for this model is

$$L(y_n | \beta_n, \alpha) = \prod_t \frac{\exp(x_{ij}(\frac{\beta_i}{\theta_i}) + z_{ij}(\frac{\alpha}{\theta_i}) + \varphi_i ASC_{ij} - c_{ij}(\frac{1}{\theta_i}))}{\sum_j \exp(x_{ij}(\frac{\beta_i}{\theta_i}) + z_{ij}(\frac{\alpha}{\theta_i}) + \varphi_i ASC_{ij} - c_{ij}(\frac{1}{\theta_i}))} \quad (2.10)$$

where x is a vector of attributes with random coefficients, z is a vector of fixed coefficient attributes, ASC is a vector of alternative-specific constants, c is the cost of the contract, and $\sigma_i = \frac{1}{\theta_i}$ is the individual-specific estimate of the scale parameter.

The model is hierarchical in nature because it builds upon the individual-specific parameters to construct the population estimators. The draws of the posterior distribution are performed by sequentially drawing from the conditional distribution of each of the parameters using Gibbs Sampling. The only computationally time consuming steps involve drawing from the posterior of the random and fixed parameters. Adding fixed variables to a Bayesian estimation adds a layer of Gibbs sampling but it was important because survey responses indicated that preferences for some attributes might not vary so much in the population. Hybrid models are rarely found in the literature, especially with non-linear specifications because of the increase in time requirements. Each of these steps involves computation of the logit formula in

Metropolis-Hastings steps. The population parameters are drawn from a normal distribution with mean at the mean of the individual-specific draws and variance W/N . W is drawn from an inverted Wishart with $K + N$ degrees of freedom, which depends on the sample variance of the individual-specific β_i s around the population mean of b .

There are several considerations with regard to the estimation of HB models. First, the researcher must specify the prior distribution of the parameters. We tested the common distributions including normal, lognormal, and truncated normal, but found that, excepting the scale parameter, the diffuse normal distribution suited our needs. Apart from the attitudinal survey distributed with the CE experiment, no prior information was available regarding values for the attributes of the contracts. While responses indicated that all attributes would be positively valued by community members, we did not want to preclude the possibility of protest votes. Therefore, to permit the possibility of positive and negative valuations, the normal distribution was chosen. It is, however, customary to specify the prior on the scale parameter to be lognormal, as, being the standard deviation of the variance of the error, it cannot logically be negative.

Balcombe, Chalak, and Frasier [92] have tested both random-walk and importance sampling M-H algorithms with varying degrees of success in obtaining convergence. Allenby and Rossi [99] point out that the random-walk algorithm works well with short panels-relatively few observations per person. We followed Train and Sonnier [100] and implemented the random-walk algorithm. The second consideration in estimation regards the details of the mechanics of the estimation. The Gibbs sampler converges to draws of the posterior distribution with enough iterations. It is common practice to discard the initial draws as ‘burn-in’ [101]. In addition, each step of the Gibbs sampler relies on the previous draw. To reduce correlation between draws, every i th

draw is retained while the rest are discarded. This practice is called ‘thinning’ the sample. The final choice to make with regard to the mechanics of the HB procedure is to determine the number of iterations to keep after convergence is assumed. The above decisions were made based on several convergence tests.⁴

While Bayesian applications are becoming more and more commonplace, social science literature remains dominated by Classical analysis. The Bernstein-von Mises theorem states that the mean of the posterior distribution is asymptotically equivalent to the maximum likelihood estimator[98]. We rely on the Bernstein-vonMises theorem when classical statistics prove informative and generally more familiar to the reader. When appropriate, we use Bayesian conventions to elaborate on our results, leveraging the richness of the HB output. Model comparison is driven by root likelihood comparisons. Root likelihood (RLH) is the geometric mean of the predicted probabilities and thus captures model fit[102]. It provides a measure of how well the model predicts in-sample behavior and is evaluated based on the number of alternatives faced by the respondents. That is, chance would correctly predict choice in a data set with four alternatives 25% of the time. A Root likelihood value of 0.544 (Table 2.3, Model 1) thus reveals that the model performs a little over twice as well as pure chance.

The code used for the hierarchical Bayes models builds upon Matlab code made publicly available on Kenneth Train’s website for the estimation of hierarchical Bayesian discrete choice models in preference space[100]. All models are implemented using the R statistical package. The output of the analysis is passed to the BOA package in R for analysis of the MCMC chains.⁵

⁴Convergence diagnostics are listed in Appendix B.

⁵ WTP space models estimated via classical procedures are most commonly implemented in BIOGEME with the CFSQP algorithm to ensure a global maximum.

2.5 Results

The contract attributes in this survey may all be considered “goods” and positively related to the health of the Bobolink species and/or consistent with supporting farmland in the community. Survey responses indicated that both of these objectives were important to respondents. We therefore expected that individual-specific parameter estimates as well as the population-level estimates of the farm-wildlife contract attributes would be highly correlated and positive.

Across the three models, all parameter estimates have the expected signs (Table 2.3). The mean willingness-to-pay per acre of protected Bobolink habitat was higher for models in which the parameter was permitted to vary across respondents. This is likely due to the influence of the Both alternative on overall values because the population-level estimates are derived from the individual level of the hierarchy of estimation. Individuals who choose both contracts for all survey questions signal that there is no cost within the tested range that would induce the respondent to make trade-offs. Therefore, any trade-off can be inferred.

Taste heterogeneity is evident for the Acres and Restore attributes of the farm-wildlife contracts but not for the View and Tour attributes, as evidenced by highly significant variance estimates. The variance parameters for both View and Tour are statistically insignificant and were specified fixed in both Models 2 and 3. The ASCs are negative indicating that choosing either the No Buy option or the Both option is not a dominant behavior. Approximately 11% of the population had positive values for the No ASC in Model 1. In fact, there were a total of 454 choice instances out of 3723, that is 12.2% of all choices, in which the No-Buy option was chosen. This indicates that the model is capturing the data generating process well.

The scale coefficient is stable across model specifications as is its estimated variance. The estimated scale variance is highly significant in Models 2 and 3 lending support to the notion that estimating a model with fixed scale would likely be a misspecification. While the scale heterogeneity model (Model 2) outperforms against the full taste heterogeneity model (Model 1), we find that incorporating some taste heterogeneity provides the best posterior fit in terms of RLH (Model 3). Model 3 permits some variety in tastes while accommodating individual-specific scale differences. The remainder of the analysis utilizes this model to explore the relationship between scale, response strategies, and individual-level model fit.

Distributions of Willingness to Pay

A comparison of the posterior distribution of the population-level parameters against the distribution of the individual-specific parameter estimates reveals the influence of yea-saying on population-level, or hyper-parameters. Here we define yea-saying as choosing both contracts for all choice tasks. Figure 2.1 plots the kernel smoothing density of the population-level estimates of the random parameters (on the left) and the kernel density of the distribution of the estimated individual-specific coefficients (on the right). Several important features of the data emerge when the distributions of the population-level are compared against the distribution of the unit-level parameters. First, one indicator of model convergence is smoothly shaped hyper-parameters (population-level parameters). Non-convergent distributions have poorly defined shapes with several peaks and troughs. Our population-level distributions have smooth shapes with well-defined maxima indicating likely convergence of the MCMC chains. The distribution of individual-level Both ASC parameters has three local maxima. Upon inspection of the data, the high-valued peak of the individual-level Both ASC is made up solely of yea-sayers. These individuals are likely pulling up the population estimate of the Both ASC. This was evident with

the Restore attribute as well. There were 251 individuals with positive values for the Both ASC. These were individuals who chose the Both option with a high probability. There were thirty individuals with positive coefficients on the No ASC. These individuals chose the ‘No-buy’ option for all or most choice occasions. The distribution of the individual-level Scale parameters is highly skewed toward zero implying that the majority respondent had values below 0.05 with outliers tending to have large scale parameters. There were thirty-four individuals with negative WTP estimates for the Acres contract attribute with an average willingness to accept value of \$0.12 per acre. These individuals were represented in the left tail of the distribution of the individual-level WTP parameter.

Root Likelihood

Because the scale parameter is linked to the variance of the error term, we examine the relationship between model fit and scale in order to determine if there are important patterns that may reveal response strategies. On average, all individuals who were identified as having high scale also had high root likelihood values. High scale implies that the variance of the idiosyncratic error term for the individual is low. If we assume that error variance is large (scale parameter is small) because of inconsistent behavior on the part of the respondent, then it makes good intuitive sense that the model will “fit” better the data in cases where the subjects make consistent, predictable choices. Another way to say this is that high scale strengthens the effect of the observed attributes on choices.

The high-scale respondents had, on average, high RLH values with a mean RLH of 0.69. The low scale respondents had both high and low RLH values. Taken together, there is no statistical difference in RLH between the highest and lowest quantiles of scale (0.69 vs. 0.64, $p=0.054$). However, if the yea-sayers are removed from the group of low scale respondents, mean RLH of the lowest quantile is halved (0.64 to

0.32) and the difference between high and low scale respondents is significant ($p < 0.000$). This implies that the model predicts the lowest scale individuals (excluding yea-sayers) only 1.28 times the rate of chance (0.25 for a four alternative choice situation), as opposed to 2.76 times better than a model of chance for the high-scale respondents. The average RLH for the yea-sayers alone was 0.90, exemplifying an important point about RLH as a measure of goodness of fit. It is this consistency in response that leads to high fitting (high RLH values) for these individuals. The Both alternative contains the attribute levels of the first and second alternatives combined, so choosing it consistently does not reveal much about preferences for specific attributes. This pattern of response leads to low scale parameters and overall high values for WTP. The low-scale respondents who did not fall into the yea-sayer category had high WTP values, but, did not always choose the Both alternative. Overall, high scale implies that more of the value of the WTP estimate translates to the value of alternative, meaning that the included variables explain values well and the RLH for the respondent is high. Low scale implies that the variables in the model do a poor job of explaining contract values and RLH for the respondent is expected to be low. Yea-sayers provide the exception to this rule because their consistency in response yields high RLH values, even though scale is low.

In order to elucidate the effects of choosing the Both alternative on the scale of responses, we focus on the lower and upper quantiles of the distribution of the individual-specific scale parameter. There were 198 individuals in each of the lowest (high error variance) and highest quantiles (low error variance). We constructed individual-level willingness to pay for a 10-acres plot with 10 acres of restored farmland, a view from the road and an invitation to an expert-led birdwalk. The individual-level estimates of willingness to pay are highly negatively correlated with scale (corr=-0.85, t=-46.04, df = 789, p=0.000). Figure 2.2 plots the Kernel density

of the individual-level parameter estimates of WTP, scale, and the ASCs by the following groupings: lowest quantile scale, highest quantile scale, and all respondents. The figures highlight the high correlation between choice of the Both alternative and the value of the scale parameter. The lowest quantile of the scale parameter distribution included all 109 individuals who chose the Both Alternative for all choice situations. Individuals who fall into the highest quantile for scale had relatively low WTP estimates and highly negative Both ASC parameter estimates. Here we see that individuals in the highest quantile of scale avoided the Both alternative. These results emphasize the importance of controlling for the effect of the choice of both contracts on the scale of the error term for individuals.

Factors that Influence Scale

In order to gain some insight into what demographic and survey-related variables beyond yeah-saying affect the value of the individual-level scale parameter, we implement a classical linear regression modelling the log of the scale parameter as a function of three types of regressors. The scale parameter was specified to follow a lognormal distribution in the first-stage Mixed Logit specification. It was therefore necessary to log-transform the variable to restore the underlying normal distribution.

The first set of explanatory variables were included to capture the effect of choosing the Both alternative in the choice situations. An indicator was included to identify respondents who chose only the Both alternative for all choice instances (Yea-say). A second indicator variable identified respondents who never chose the Both alternative (No Both). Finally, the number of choices for which the Both alternative was chosen was included (Number of Both Alts).

The second set of regressors included a variety of demographic attributes supplied by the marketing firm, such as gender, income, donation and mail-order history. Also included were two variables constructed from the factor analysis of the attitudinal

section of the survey. The Attach co-variate captured the respondent's level of attachment to the Jamestown community. The Equality variable represents the second factor in the factor analysis of the survey and indicates the degree of interest in guaranteeing equal payments for all residents.

The final set of regressors represented the treatment groups representing the elicitation mechanisms. There were eight separate treatment groups. The two main groups differed in whether the reference mechanism was supplied for the first two questions and each was further subdivided by elicitation mechanism, provision point with money-back guarantee (PPMBG), proportional rebate (PR), uniform price auction (UPA), or pivotal mechanism (PM). The best performing models are listed in Table 2.4.

At the outset of the analysis, it is important to remind readers that there is an inverse relationship between the estimated scale parameters and the magnitude of error variance for the respondent. That is, higher scale parameters imply lower error variance and thus, more consistent preferences across choice situations. The effect of yeah-saying on error variance is significant and highly negative across models. This result highlights an important drawback of allowing individuals to choose, essentially, not to choose. In that case, it is essentially impossible to construct estimates of the necessary trade-offs between attributes that help define the relative values of the attributes. The Number of Both Alts and No Both variables are highly significant and their effect is robust across model specifications. As the number of Both alternatives chosen rises, the value of the estimated scale parameter falls. Similarly, individuals who never indicated that they would choose both contracts had significantly higher scale parameters. By controlling for these effects, we can analyze the impacts of demographic characteristics and treatment on choice consistency.

Several demographic characteristics influence choice consistency. Perhaps not surprisingly, a history of donations to environmental causes or organizations raises the consistency of choices as does having a history of donating to health-related causes. In addition, having some history of ordering items through the mail raises the consistency of choices. It may be that these individuals are more accustomed to spending time and attention on items received in the mail.

Of the factor analysis variables, indicating a high level of attachment to the Jamestown community in other sections of the survey did not have implications for the consistency of choice of the respondent. On the other hand, the Equality variable was found to be positively related to the scale parameter implying that an interest in equal payments by all community members led to more consistent choices. The reason for this link is not evident.

Scale and The Rules of the Game

Finally, there are important relationships between the mechanism treatments and the value of the scale parameter. Recall that the surveys were divided into two groups. The first (Group 1) was administered a reference mechanism: a hypothetical referendum for a tax increase in the amount of the cost of the contract. The second group (Group 2) received only their assigned elicitation mechanism: PR, UPA, or PM. There were approximately twice as many respondents in the treatment that received the reference mechanism (Group 1) than in the treatment that received none (Table 2.5). Although there are thirty individuals relegated to the Group 2 PPMBG sub-group, this set of individuals in actuality were administered Group 1 PPMBG surveys but did not complete the referendum questions.

The base treatment variable for this analysis was the pivotal mechanism without the reference mechanism (Group 2). An F-test for the joint significance of the Group 2 treatments, suggests that there is no difference between the consistency of response

across mechanism treatments in this group ($p = 0.66$). Relative to the base treatment, all Group 1 treatment sub-groups had lower scale parameters. This follows naturally from the fact that the switch from reference mechanism to the alternative auction mechanism induced a change in valuation that was not captured in the model. Had the reference mechanism treatments failed to significantly affect change, we would assume that there was no change in behavior across choices and might hypothesize that the respondents were somewhat insensitive to the shift. This provides evidence to support our first hypothesis. Individuals seem to be changing their behavior when the ‘rules of the game’ change.

Within the group that did not receive the referendum questions, the PP/MBG has markedly low scale parameters. However, as mentioned above, this group likely included the individuals who, for some reason did not complete the referendum portion of the choice tasks. Thus, low scale parameters may simply imply that there is less preference information as a result of not completing the entire experiment.

According to the results of the regression model, the ordering of mechanisms by choice consistency is as follows:

$$PM = PR > UPA > SPUPA > SPPM > SPPR > SPPP/MBG > PP/MBG \quad (2.11)$$

The proportional rebate mechanism does not perform significantly different from the pivotal mechanism. At first glance, it appears that the pivotal mechanism is not in fact being hindered by its complexity (Hypothesis 2) as, in both main groups, its scale parameter is relatively high. In addition, there is some weak evidence that the uniform price auction yields lower scale parameters than the proportional rebate and pivotal mechanisms. At the very least, the results imply that the UPA performs no better than other mechanisms at achieving choice consistency. This provides some basis for rejection of our third hypothesis. That is, there does not seem to be gains

in choice consistency to be had by couching the exchange in a manner that is more comparable to a private market transaction. However, as mentioned above, we note that we did not find a significant difference in scale within the two main groups.

In order to evaluate whether the different elicitation mechanisms have an effect on free-riding behavior, we use individual-level willingness-to-pay values (Table 2.6). The average WTP was higher for the Group 1 respondents. This finding does not necessarily indicate that in making the respondent more aware of the mechanism, it is having the intended impact of reducing free-riding (thereby inducing the respondent to reveal their true, higher values). It may simply be the case that the referendum questions are pulling up the estimates of WTP.

There was no significant difference in either scale or WTP across elicitation mechanisms within each main group. However, while not significant, we note that, of the Group 2 treatments, the pivotal mechanism had the lowest contract WTP. Thus, there is weak evidence that the PM mechanism is failing to induce respondents to reveal their true valuations when a reference mechanism is not administered. However, by comparing each mechanism individually across Groups 1 and 2, we get a sense of just how much changing the payment rules makes individuals more responsive to the mechanism. For instance, we cannot reject the null hypothesis of equality of means in scale parameter across groups for individuals who were administered the UPA mechanism and the PP/MBG mechanism. For these mechanisms, applying a reference mechanism did not induce a change in behavior, presumably because the “rules of the game” did not change sufficiently to warrant a change in strategy. There was, however, a marked difference between groups for both the PR and PM mechanisms. The PM, the only incentive-compatible mechanism tested in this experiment, had the highest mean scale parameter when administered without a reference mechanism (0.064). The difference across the two groups for the PM was 0.02, 2.5 times the

difference across groups for the PR, the only other statistically significant difference (0.008). This finding suggests that much of the difference in scale between the two groups is driven by the difference in the PM vs. SP/PM. That is, individuals appear to be reacting more to the change in mechanism from referendum to PM than to other mechanisms.

The difference in average WTP across the two groups by mechanism is even more revealing (Table 2.6). The pivotal mechanism is the only mechanism with a significant difference in WTP across the two groups. In addition, the PM produced the lowest values for WTP in the Group 2 treatment, but, highest the values in the Group 1 treatment. We can, therefore, comfortably conclude that the increase in WTP may be due to respondents reacting to the mechanism by more truthfully revealing their values. Thus, providing a reference mechanism may have the effect of inducing the respondent to behave in a manner more aligned with behavior in a laboratory setting.

2.6 Summary and Conclusions

Constructing a market for ecosystem services involves designing a market for public goods which suffer from free-riding. Generally, the first step in designing markets for new products involves surveying respondents about their preferences for the attributes of the new product. The Jamestown choice experiment was designed to test different payment elicitation mechanisms and their ability to mitigate free-riding behavior in a local market for wildlife habitat protection.

Because there is no real-world cost of their choice, respondents may be motivated to overstate their WTP for new products. The Jamestown choice experiment provides a straightforward means to assess the impact of this type of response strategy on WTP calculations from hypothetical choice experiments. We explored how yeah-saying affects attribute valuation as well as scale. For our purposes, yea-saying manifests as consistently choosing the Both Alternative for all choices presented to the respondent.

By employing a random-parameter logit model in WTP-space, we were able to examine the impact on scale and marginal values of the yea-saying strategy. We found that yea-saying individuals had yields low scale parameter estimates, high WTP values, and high fit. High attribute values likely have a strong upward pull on hyperparameters in the model. Therefore, inferences based on the population estimates will be biased upward. Yea-sayers were likely to have high root likelihood values, indicating consistency in choice, however, low scale implies that there is less explanatory power in the marginal WTP estimates.

In addition, we were able to explore response to elicitation mechanisms by individual. The individual-level parameter estimates of scale indicated that applying a reference mechanism induced a significant reaction to the less familiar test mechanisms with the largest response being to the more complicated but incentive compatible pivotal mechanism. To our knowledge, this is the first research to test elicitation mechanisms outside the lab and these preliminary findings indicate that elicitation mechanisms have an impact on response and that applying a reference mechanism may enhance their effects.

Further research is needed in testing elicitation mechanisms in the field in order to make conclusive statements about their effectiveness in mitigating freeriding outside the laboratory. In fact, we acknowledge that, because the good for offer was not before offered to the community, it is difficult to imagine what unobserved covariates might come in to play in the choice experiment. Because scale heterogeneity measures the variability of unobserved influences on choices, it may be possible that other explanations for choice consistency arise. It would have been helpful to interview respondents subsequent to survey completion, but, thus is the drawback of field research. However, this research provides strong clues regarding the ways in which public goods auction mechanisms behave in field experiments. This is an important and ongoing source of

research, particularly with regard to the design and implementation of local markets for ecosystem services. With the growing interest in local solutions to depletion of ecosystem services, this research provides an example of what may and may not work.

Table 2.1. Summary of contract attributes

Attribute	Description	Levels
Acres	Number of acres to be placed under contract upon which farmer will delay mowing and harvesting	10, 25, 40, 55
Restore	Number of acres to be restored to active hay fields, not restricted to delayed mowing/harvesting	0, 10, 20, 30
High Bobolink	Level of expected fledglings saved (correlated with acreage)	Low, High
View	Whether the parcel will be viewable from the road	View, No View
Tour	Whether individuals paying into the contract are invited to an expert-led birdwalk	Tour, No tour
Cost	Cost of the Contract	\$10, \$20, \$35, \$45, \$60, \$75, \$85, \$105

Note: The five questions from the CE experiment included in this analysis presented two contracts side by side with varying levels of the above attributes. Participants were asked to choose their preferred contract, both, or none. High Bobolink was excluded from the analysis because we were not able to identify its influence independently of the Acres attribute. Source: [76]

Table 2.2. Sources of Heterogeneity in Choice Models

Specification	Implications
Fixed ASCs	differences in residual tastes across alternatives, equal across respondents, i.e. “brand loyalty”
Random ASCs	differences in residual tastes across alternatives that varies by respondent type, residual taste heterogeneity
Fixed scale	heterogeneity in the variance of the error term across choices but assumed equal across individuals
Random scale	heteroskedastic error variance, scale heterogeneity
Fixed attribute parameters	heterogeneity in preferences for attributes
Random attribute parameters	heterogeneity in preferences for attributes that varies by type of respondent

Note: The table above outlines the different possible model specifications when modeling choice experiment outcomes and describes their implications. Source: Author.

Table 2.3. Model Comparison

Attribute	Model 1	Model 2	Model 3
	Parameters		
WTP Acres	0.79 (0.09)	0.51 (0.05)	0.75 (0.10)
WTP Restore	1.20 (0.13)	0.82 (0.08)	1.29 (0.13)
WTP View	10.85 (2.00)	8.31 (1.42)	8.41 (1.45)
WTP Tour	10.31 (2.12)	5.01 (1.58)	5.08 (1.61)
No ASC	-2.63 (0.24)	-4.21 (0.34)	-3.66 (0.37)
Both ASC	-1.36 (0.22)	-0.77 (0.20)	-1.13 (0.24)
Scale	0.024 (0.001)	0.028 (0.002)	0.030 (0.002)
	Variance		
WTP Acres	0.39 (0.14)		0.38 (0.10)
WTP Restore	0.44 (0.20)		0.72 (0.22)
WTP View	2.91 (3.77)		
WTP Tour	6.36 (6.04)		
No ASC	5.70 (1.22)	13.25 (2.27)	7.09 (2.30)
Both ASC	15.10 (2.19)	10.76 (1.50)	10.53 (1.47)
Scale		0.81 (0.14)	1.07 (0.19)
LL	-2816	-2468	-2323
RLH	0.516	0.563	0.586
ParamRMS	6.47	3.76	2.67
Avg. Variance	38.67	15.08	7.58

Note: Standard Errors are in parentheses.

Table 2.4: Scale Regression

VARIABLES	Model 1 lnprice	Model 2 lnprice	Model 3 lnprice	Model 4 lnprice
Survey Design Related Attributes				
Yeah-sayers	-0.223** (0.0937)	-0.286*** (0.0934)	-0.238** (0.0960)	-0.282*** (0.0915)
Number of Both Alts	-0.186*** (0.0258)	-0.164*** (0.0255)	-0.178*** (0.0264)	-0.169*** (0.0250)
No Both	0.461*** (0.0627)	0.556*** (0.0633)	0.532*** (0.0637)	0.489*** (0.0623)
Treatment Mechanisms				
PPMBG	-0.492*** (0.114)		-0.481*** (0.116)	
PR	-0.102 (0.0829)		-0.113 (0.0847)	
UPA	-0.136 (0.0843)		-0.130 (0.0860)	
SP-PPMBG	-0.295*** (0.0744)		-0.315*** (0.0760)	
SP-PR	-0.243*** (0.0726)		-0.276*** (0.0743)	
SP-UPA	-0.149** (0.0731)		-0.178** (0.0748)	
SP-PM	-0.193*** (0.0743)		-0.223*** (0.0755)	
Respondent Demographics				
Attached	-0.0155 (0.0189)			-0.0242 (0.0190)
Equality	0.101*** (0.0193)			0.104*** (0.0195)
Income	-0.000241 (0.000245)			-0.000248 (0.000247)
Purchasing Power	-0.000441 (0.000346)			-0.000365 (0.000349)
Donations, Any	-0.152** (0.0685)			-0.148** (0.0694)
Donations, Political	0.0652 (0.0843)			0.0580 (0.0852)

Continued on the next page

VARIABLES	Model 1 lnprice	Model 2 lnprice	Model 3 lnprice	Model 4 lnprice
Donations, Environmental	0.180** (0.0712)			0.183** (0.0719)
Donations, Health	0.249*** (0.0669)			0.235*** (0.0678)
Donations, Religious	-0.190* (0.113)			-0.154 (0.114)
Mail Order, Electronics	-0.0903 (0.0608)			-0.111* (0.0615)
Mail Order, Magazines	-0.114 (0.106)			-0.106 (0.107)
Mail Order, Kids Magazines	0.159 (0.109)			0.148 (0.110)
Mail Order, Any	0.100*** (0.0385)			0.114*** (0.0389)
Head of House	0.0338 (0.0410)			0.0195 (0.0412)
Constant	-3.122*** (0.0934)	-3.351*** (0.0579)	-3.118*** (0.0813)	-3.339*** (0.0748)
Observations	791	791	791	791
R-squared	0.619	0.569	0.587	0.603

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.5. Summary of treatments administered to respondents

Mechanism	Group 1	Group 2
Provision point, MBG	126	30
Proportional Rebate	144	76
Uniform Price Auction	139	72
Pivotal Mechanism	131	73
Total	540	251

Note: Group 1 received the hypothetical referendum base mechanism, group 2 did not. In addition, the Provision Point, MBG was not administered to group 2. The 30 individuals listed in the Group 2 PPMBG category were administered Group 1 PPMBG surveys but failed to complete the hypothetical referendum questions.

Table 2.6. Tests of Differences In Means by Mechanism Treatment

	Group 1	Group 2	H0: Equality of Estimates
Scale			
Pivotal Mechanism	0.046	0.065	Reject (p=0.029)
Proportional Rebate	0.045	0.056	Reject (p=0.026)
Uniform Price Auction	0.051	0.061	Cannot Reject (p=0.304)
Provision Point, MBG	0.046	—	NA
			H0: Equality of Means
WTP			
All	34.65	32.77	Reject (p=0.0051)
Pivotal Mechanism	35.21	31.86	Reject (p=0.0085)
Proportional Rebate	34.67	33.42	Cannot Reject (p=0.3339)
Uniform Price Auction	34.21	32.57	Cannot Reject (p=0.2125)
Provision Point, MBG	34.46	—	NA

Note: Group 1 received the hypothetical referendum base mechanism, group 2 did not. In addition, the Provision Point, MBG was not administered to group 2. The hypothesis tests of equality of scale parameters across treatments are based on the parameter estimates from the best performing models from the scale regression. The WTP comparisons are based on equality of means tests.

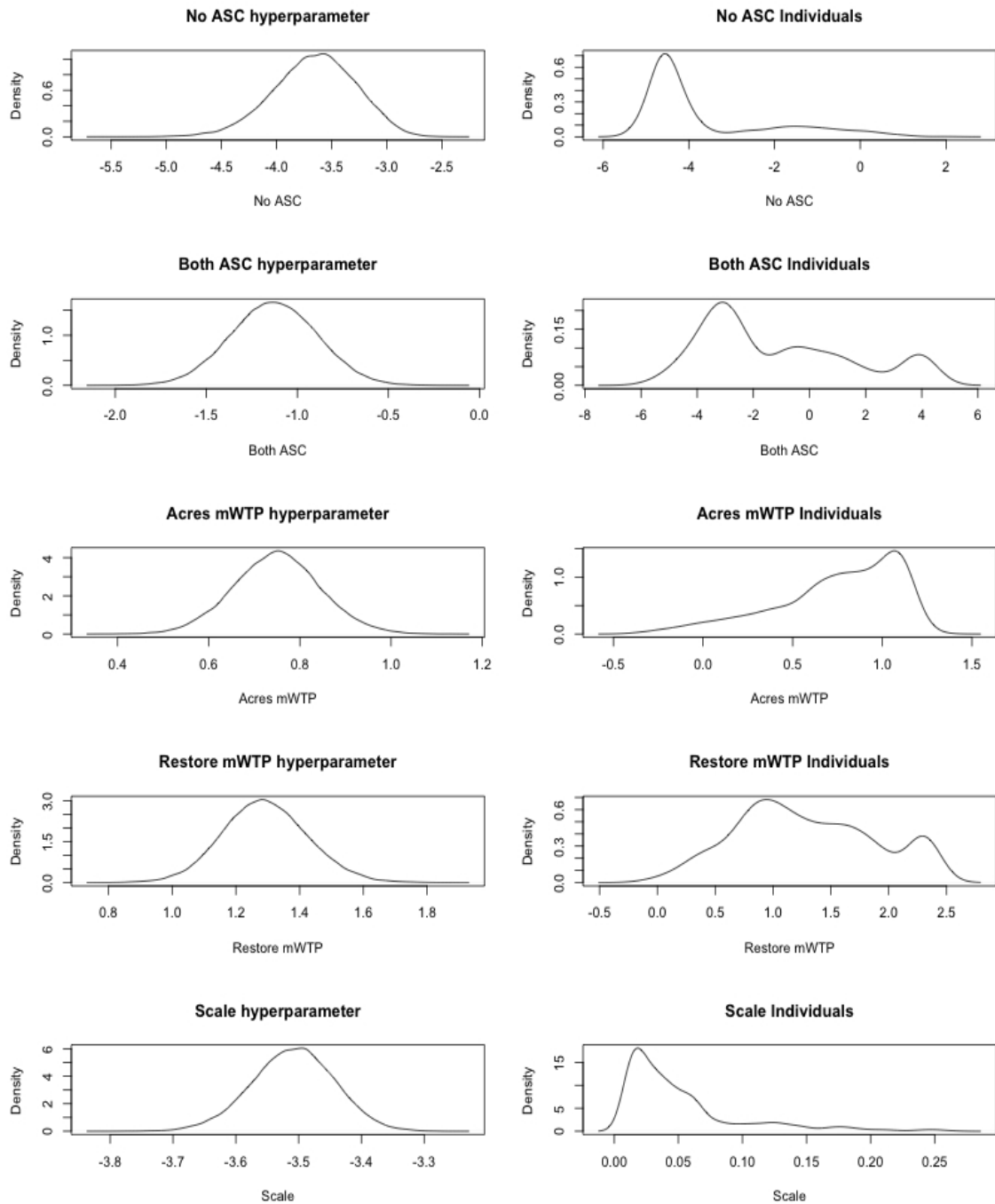


Figure 2.1. Posterior Distribution of Population Parameters and Distribution of Individual-Specific Parameters in the sample

Note: The left hand plots represent the kernel densities of the hyper-parameter, or population-level parameters. The right-hand plots represent the kernel density plots of the individual-specific estimates of the parameters. Model convergence requires that the hyper-parameters have smooth shapes with well-defined maxima.

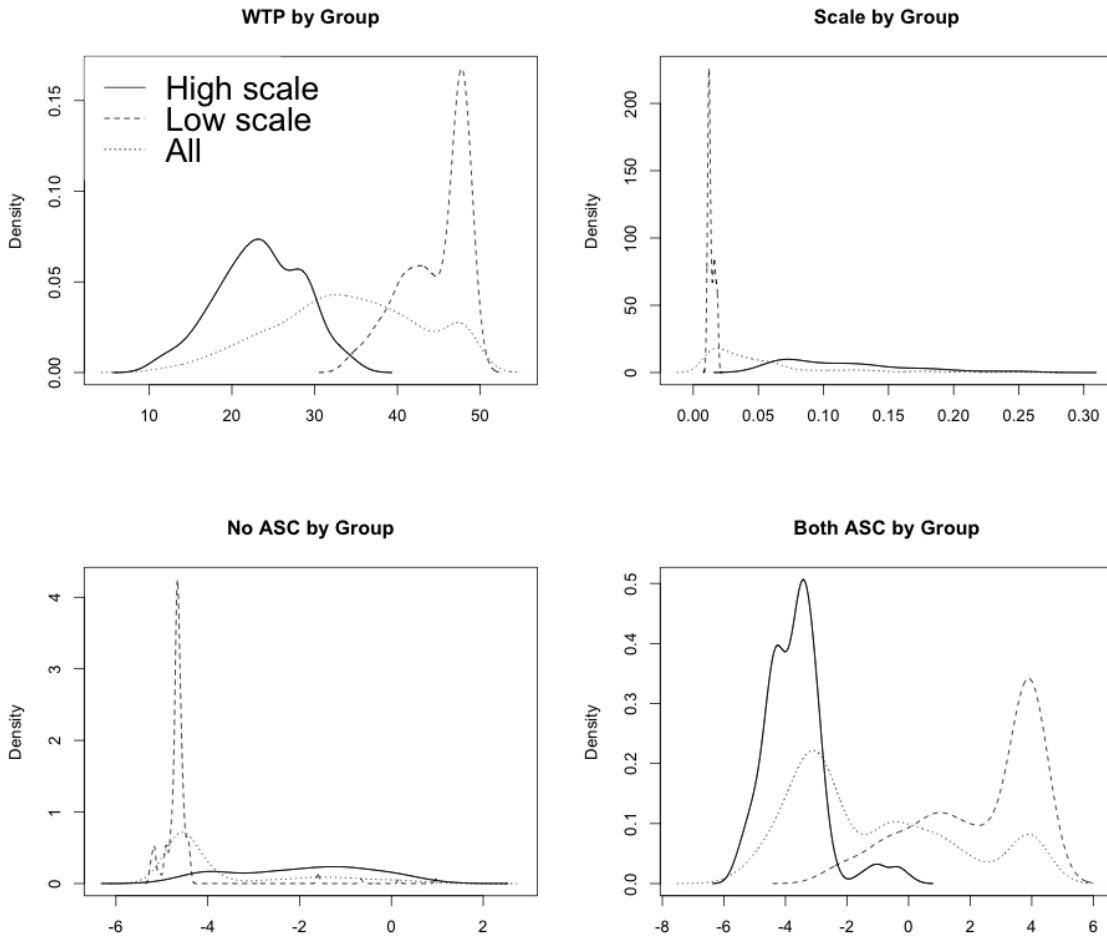


Figure 2.2. Distribution in Parameters across Extremes of Scale

Note: The figures above plot the kernel densities of the individuals with estimates of scale parameter that fall in the highest and lowest quantiles of the distribution, as well as the overall distribution. Respondents were categorized as having *High Scale* if their estimated scale parameter exceeded 0.06187 or *Low Scale* if their scale parameter fell below 0.01926.

**Predictive Validity in Valuation Assessment For the Provision of
Farmland Amenities: A Disaggregated Approach**

*Prepared for submission to the Journal of Environmental Economics and
Management*

Julie Insignares Santos⁶

⁶PhD Candidate, Department of Environmental and Natural Resource Economics, University of Rhode Island, Kingston, RI 02881. Email: jinsignares@my.uri.edu

Predictive Validity in Value Assessment For the Provision of Farmland Amenities: A Disaggregated Approach

3.1 Introduction

Research in environmental and natural resource valuation rely heavily on hypothetical survey data to estimate values for public goods [103, 104, 105, 106, 107]. The advantage of the stated preference approach lies in the ability to construct hypothetical scenarios in which the researcher has the ability to define the attributes of the scenarios and their levels [108]. In contrast, revealed preference experiments confine the researcher to the realm of realized behavior.

Often, however, results from hypothetical surveys have not been satisfactorily reflective of observed behavior. Data from stated preference experiments tend to overestimate actual demand, particularly in the case of public goods valuation [109, 110, 111, 112, 113]. This phenomenon is generally termed “hypothetical bias” and has been linked to: subject pool variety, differences in information provided across experiments, social norms, and whether willingness-to-pay vs. willingness-to-accept is being measured [114]. However, in light of the fact that market-based instruments for valuing ecosystem services are on the rise [115, 116, 117, 118, 119], demand for the kind of information that stated preference methods alone can offer is steadily rising. Inferring values from stated preference surveys that suffer from hypothetical bias may induce policy-makers to set policy objectives at levels that will result in inefficient outcomes.

Several studies have found significant differences between stated and revealed values for goods and services that are derived from the environment. Aadland and Caplan [120] compared stated and revealed preferences for curbside recycling pro-

grams. Brooks and Lusk [121] compared survey responses to scanner data on sales of organic and rBST-free milk. Champ and Bishop [122] utilized certainty scales to identify hypothetical bias in response to questions regarding the voluntary purchase of wind-generated energy for a period of one year. Murphy and colleagues [111] evaluated hypothetical payments and binding offers to contribute voluntarily to the Massachusetts Chapter of The Nature Conservancy using a cheap-talk script to mitigate hypothetical bias.

Concurrently, there is mounting evidence of the importance of accommodating attribute processing heterogeneity such as ignoring or ‘non-attendance’ to one or more features of the good and lexicographic preferences in stated and revealed preference data analysis [123, 124, 125, 126, 127, 107]. The recent literature recognizes that individuals responding to surveys often employ simplifying strategies that are not consistent with conventional random utility maximization. I posit that attribute processing rules may explain some of the observed hypothetical bias in data from surveys concerning values for public goods. The objective of this study is to identify differences between stated and revealed payments by drawing upon the latest research on attribute processing rules (APRs) in order to examine the extent to which APRs can be useful for identifying sources of hypothetical bias.

From a disaggregated perspective, I identify response strategies and test their impacts on estimated values for agricultural ecosystem services. Using a latent class model that incorporates APRs, I am able to identify strategies in responses such as ‘yea-saying’ [114], attribute non-attendance (ANA) [123], and lexicography on particular attributes [127] in order to make inferences about how these dynamics affect measurements at the aggregate level. These issues are particularly relevant to valuation of ecosystem services for several reasons. First, because of the non-rival nature of public goods and services, as long as enough support is generated, non-payers cannot

be precluded from consumption of the good. Thus, in a hypothetical situation, it is advantageous to send a positive signal. In addition, because such goods are not traded in conventional markets, assigning economic values for them can prove cognitively arduous. Thus, employing simplifying heuristics when making choices might be more likely than for private goods for which monetary valuation comes more naturally. I contribute to the growing literature on the importance of accommodating different attribute processing rules by incorporating process heterogeneity in the analysis of a choice experiment regarding farmland ecosystem services and testing the validity of the results against revealed market demand. To my knowledge, this is the first work to examine attribute processing strategies in hypothetical markets for public goods and their implications for realized behavior in the field.

In contrast to studies that utilize survey measures to identify and calibrate hypothetical responses (i.e. cheap-talk scenarios and certainty scales), the approach used in this analysis can be used in the absence of such measures. I utilize stated preference choice experiment (CE) and revealed preference market experiment (RP) data from a project involving the design and implementation of a local market-like process for ecosystem services in the community of Jamestown, RI. This particular experiment was uniquely conducive to analysis at the level of the individual for several reasons. Both CE and market experiments were administered to the same sample of individuals in a small community in rural Rhode Island. Thus, differences in subject pool (i.e. university students vs. grocery store shoppers) can be ruled out. In addition, a substantial amount of demographic and attitudinal information was collected regarding the respondents. This information is used to make assessments of feasible processing strategies. In contrast to most studies using stated and revealed preference data, this project conducted a choice experiment before the market experiment. Finally, care was taken to engineer the market good and the mechanism of exchange

to be as closely consistent with the hypothetical choices as possible. Moreover, nearly identical elicitation mechanisms including the pivotal mechanism and provision point mechanisms were administered in order to reduce freeriding. This way, I am able to examine the performance of different elicitation mechanisms to address free-ridership as a separate issue.

The paper proceeds as follows. Section 3.2 describes the ecosystem service under analysis and provides a brief review of the CE and RP experiments and the hypotheses to be tested. Section 3.3 describes the method used to combine the data. Section 3.4 presents results and section 3.5 provides a summary and conclusion.

3.2 Constructing a local market for wildlife preservation

The empirical application involved establishing a market for wildlife protection marketable to the members of the rural communities surrounding local farmland. Every spring, hay farms on Jamestown, Rhode Island serve as nesting grounds for a species of ground nesting birds with a large migratory range and charismatic song called the Bobolink (*Dolichonyx oryzivorus*). Historically, hay fields in many U.S. states have been in decline as preference is given to other crops. In addition, crops are cut 2-3 weeks earlier than has been historically (since the 1940s and 1950s) [128]. This shift in cropping practices has led to serious mortality for Bobolink fledglings. Consequently, Bobolinks are now protected under the Migratory Bird Treaty Act and are listed as a Species of Special Concern in some states. The Bobolink experiment sought to transfer compensatory payments from community members to farmers to delay harvesting of hay crops thereby permitting Bobolink offspring to fledge and avoid devastation at the hands of the plow.

The two phases of the experiment were administered as a mail-in survey and solicitation to the inhabitants of Jamestown, RI and spanned a three-year period from 2006 to 2008. In all treatments, the target mailing was all deliverable addresses in

the town. The SP survey was designed as a multi-question choice experiment (CE) experiment mailed to the residents of Jamestown from October to December of 2006. There were 5 questions comparing two potential contracts and a sixth question with one potential contract. The sixth response was not utilized in this analysis. Each contract was described by a list of attributes (see Table 3.1). There were six attributes described: 1. acreage under contract to delay harvest (Acres), 2. the number of acreage to restore to active farmland (Restore), 3. whether the acreage was found to have a high or low concentration of Bobolink (HighBobolink), 4. whether the contracted acreage is viewable from the road (View), 5. whether or not a birdwalk is offered (Tour), and 6. the cost of implementing the contract. Respondents were presented with two competing contracts displayed side by side. Individuals were then asked whether they would choose contract A, contract B, both, or neither. That individuals were given a choice to choose Both contracts is a novel feature of the SP survey that permits identification of yea-sayers in the sample. A full description of the survey design and implementation can be found in Uchida et al.[76].

The SP mailing was comprised of five sections in total. In addition to the CE task outlined above, there were three additional sections that elicited opinions with respect to values of farmland amenities, rural character preservation, community attachment, and the importance of fairness in payment for services provided by farmland amenities. This last line of questioning was meant to assess the impact of different elicitation mechanisms on the decision to participate in the market. Several public goods payment mechanisms were administered in order to test in their field effectiveness at mitigating free-riding. A summary of the relevant findings of the attitudinal sections is listed in Table 3.2. The last section of the mailing collected demographic information from the respondent.

The second phase of the project was a revealed-choice experiment whereby acceptable contracts were drawn up between the mediators (researchers at the University of Rhode Island in association with EcoAsset Markets, Inc.) and farmers in the same community of Jamestown, RI. Community members were solicited for payment toward provision of the contracts, again via mail-in solicitation. This phase of the project, marketed as The Nature Services Exchange of Jamestown, was open to the residents of Jamestown in early 2007 and again in early 2008. The markets successfully provided five of ten potential field contracts.

The SP and RP treatments were designed to be as closely consistent as possible. However, there were some differences across surveys. A comparison of the attributes across both phases of the project is listed in Table 3.3. The market experiment did not offer participants a choice to restore fallow land to active cultivation. Also not included in the market process was an invitation to an expert-led birdwalk. Plot support for Bobolink activity was represented differently across treatments. In the CE experiment, the attribute “Low/High Bobolink Concentration” was mildly correlated with plot size and, in all specifications, was found to be insignificant. The market experiment included two separate measurements of potential to support Bobolink populations. The number of Bobolink territories observed in 2006 was used in the 2007 treatment and the number of Fledglings expected to be supported by the field was used in 2008. Both CE and market treatments described contracts as having a view or no view, with the addition of a ‘partial view’ option in the 2007 market treatment. The field size attribute had a broader range in the stated preference survey and costs were comparable across years.

Market participation, as expected, was higher for individuals who returned the SP survey. Of the 791 respondents who returned the SP survey, 764 and 713 respectively were included in the market mailing in 2007 and 2008. The individuals who chose

the "Both" option consistently in the SP experiment are singled out as "yea-sayers". There are two specific issues with this group. First, choosing the Both option for each question offers no information about one's relative preferences for the attributes of the contracts. Specifically, no trade-off between cost and attribute levels is observed. Thus, a model that assigns either arbitrarily low values for the marginal utility of income or arbitrarily high values for the marginal valuation of all attributes is likely to result. These respondents will likely pull the population estimates up. Of the original set of CE respondents, there were 109 (13.8%) individuals who chose the Both option for all choice situations.

The second issue with this subset of the SP respondents is the actual signal that is being sent by adopting a "yea-saying" strategy. Caudill and colleagues [109] present evidence that yea-sayers come in two varieties: that some of the respondents are truly more interested in and willing to pay for ecosystem preservation but that others may not ultimately be willing to pay the stated amount. This second group may simply be sending a signal that farmland amenities are important to them without expending the mental effort to assess whether they would actually be willing to pay the stated amount. Because of the hypothetical nature of the survey, there is no consequence to this type of behavior and this type of behavior is generally termed "hypothetical bias".

Over all, individuals who returned the CE survey were more likely to return the market experiment solicitation, make an offer, and offer higher bids. The subset of yea-sayers, in fact, have even higher participation and offers (with the exception of the 2008 treatment). However, the vast majority of respondents to the CE experiment stated that they would participate in at least one of the scenarios offered them. Given this fact, the participation rate for these individuals is lower than one would expect. There were two possible reasons for lower participation rates in the 2008 treatment

of the market experiment. First, the recession by then was fully in place and, second, there had been some controversy involving the land trust's efforts to purchase conservation easements on three farms in Jamestown. This may have had the effect of generally reducing confidence in projects to support farmland. A summary of market experiment participation separated by year and by major group (SP respondents, SP yea-sayers, and RP response only) is listed in Table 3.4. Based on this information, I test several hypotheses.

First, I exploit techniques from burgeoning research on attribute non-attendance (ANA) to test outcomes for individuals who had low sensitivity to the contract attributes that did not transfer to the market experiment. Attribute non-attendance involves ignoring one or more attributes when comparing alternatives in a choice scenario. I model attribute non-attendance to the Restore and Tour attributes and hypothesize that these estimates of stated WTP are in a sense more reliable because they assume that the decision process is more aligned with the RP scenario.

Hypothesis 1: Non-Attendance to attributes that were not included in the market experiment yields more consistent estimates of WTP.

Next, I examine the market behavior of yea-sayers in this context. Because of the unique presence of the Both option, yea-saying is easily detectable in the SP application. Based on the discussion above, I examine the behavior of this class of individuals in the market in order to determine the extent of hypothetical bias inherent in yea-saying. I then compare the incidence of hypothetical bias in this group against the rest of the sample.

Hypothesis 2: There is both higher incidence of hypothetical bias and higher revealed payments linked to yea-saying behavior.

The overarching goal of this analysis is to explore whether accounting for attribute processing rules impacts the predictive validity of SP data. While the SP survey required substantially more cognitive effort and costs in terms of investment of time, the RP survey traded these costs for actual monetary commitments. Therefore, while both experiments involved costs, they were of different types. Indeed, Hensher and Greene [126] suggest a link between attribute processing rules and hypothetical bias, implying that failure to accommodate for APRs might significantly contribute to what has been termed hypothetical bias in the literature. Thus, I examine a model of response that accounts for preference heterogeneity alone against one that incorporates attribute processing rules. I hypothesize that models that incorporate APRs outperform models based on random utility maximization in terms of both model fit and predictive validity.

Hypothesis 3: Behavioral outputs of SP measures that account for APRs predict payments in experiments involving real payments with greater accuracy than models that account for preference heterogeneity alone.

3.3 Methods

The analysis proceeds in four steps. First, two competing models of SP response are estimated: a latent class model (LC) which incorporates taste heterogeneity and a latent class model that also accommodates attribute processing rules (LC-APR). Then, in order to make inferences about how certain processing strategies manifest in revealed preference experiments, the individual-specific conditional probabilities of class membership from the LC-APR model are utilized in a model of market participation for the 2007 and 2008 market experiments. Conditional on market participation, an offer equation is used to examine differences in offer amounts by respondent type, controlling for contract and demographic covariates. Specific attention is paid to classes that model non-attendance to attributes that were omitted from the RP

experiment. I test whether there is evidence that non-attendance to the Restore and Tour attributes leads to more reliable estimates of market outcomes. Finally, predictive validity of the LC and LC-APR models is compared.

The LCL Model

In order to identify strategic behaviors that might violate the assumptions of neo-classical utility maximization and evaluate the predictive performance of a model that incorporates these strategic behaviors, I utilize an approach that is commonly used in the absence of direct survey queries about response rules. The latent class logit (LCL) model with restrictions for APRs is particularly useful for this kind of analysis. The LCL model has been used both to explore patterns of attribute non-attendance and other violations of continuous preference ordering as well as modeling non-parametric preference heterogeneity. There is by now a substantial literature which uses the LCL model to identify attribute processing rules (APR), especially in the absence of self-reported non-attendance [123, 124, 125, 126, 127]. Most cite improvements in model fit and more realistic estimates of WTP when attribute processing strategies are incorporated in this manner. The LC model is widely exploited in marketing and transportation studies but has recently been used in cases of public goods valuation. Several studies cite the importance of accommodating APRs in choice modelling and there is growing evidence that modeling APRs improves model fit and leads to estimates of marginal WTP that are more consistent [107, 28, 125, 124]. If an attribute is ignored, then relative trade-offs that involve that attribute are not meaningful. That is, no increase/decrease in the ignored attribute compensates for a change in an attended attribute. This is particularly concerning if the attribute being ignored is the Cost attribute as WTP estimates cannot be calculated.

The first-stage model combines Train's [129] Expectation Maximization algorithm for nonparametric estimation of the random parameter latent class logit model with

Hess et al.'s [130] expansion of attribute non-attendance for heterogeneous taste variation. I model non-attendance to key CE survey attributes that were omitted in the market experiment: acres of restored farmland and invitation to a bird walk. In addition, I aim to catch and contain the yea-sayers in the sample, whose insensitivity to contract price would otherwise inflate marginal values for all other attributes. Attribute non-attendance is expected to be a significant problem with this particular type of choice task since respondents are not likely to be familiar with the ecosystem service for offer and thus may make unforeseen assessments of the true meaning of the attributes of the contract, or to decide that a particular attribute is too cryptic to assess a value for.

The Expectation-Maximization algorithm applied to latent class modelling has been utilized by Train [129] as a form of non-parametric estimation of underlying taste heterogeneity whereby a discrete distribution whose accuracy in approximating the true underlying distribution rises with the number of parameters. This is an extension of Bhat [131] where increasing the number of classes allows for better approximation of taste heterogeneity. Several authors have noted advantages of the LC model over the popular Mixed Logit Model in capturing taste heterogeneity [132, 133, 130].

The latent class specification proceeds as follows. Given the standard choice modeling scenario, N agents choose among J alternatives in each of T choice occasions, let y_{njt} be an indicator variable equal to 1 if agent n chooses alternative j in choice situation t . Each alternative is defined by a set of attributes with varying levels and the choice of attributes is assumed to result from standard neoclassical random utility maximization. Further assume that there are C distinct sets of taste parameters in the population, $\beta = \{\beta_1, \dots, \beta_C\}$. A set of C Multinomial Logit models represent the C discrete support points for the distribution of tastes in the population. In this framework, there are two sets of unknowns: the β_c s, that is, the estimates of taste

parameters for each class, and the class membership status of the agents. The number of classes, C , is chosen by the researcher based on measures of fit. If agent n belongs to class c , then the probability of observing her choices is the product of logit formulas over all choice situations. A set of C conditional logit models weighted by class share are repeatedly estimated. The weights are constructed via fractional multinomial logit. One of the advantages of the LC model is the ability to model heterogeneity in preferences via the class membership model. By including respondent demographics one can model heterogeneity in preferences without the need to interact demographics with the attributes of the choice situation. The class membership model is estimated simultaneously via fractional multinomial logit.

Model fit is generally assessed based on minimizing an information criterion such as AIC, BIC, or CAIC [123, 107, 134]. If the information criteria do not agree on which model is preferred, the researcher must choose based on examination of standard errors and feasibility of parameter signs. Because the EM algorithm does not involve maximizing the likelihood function, special attention must be paid to assessing local vs. global maximum attainment. This is achieved by testing several starting points to ensure that a global maximum has been obtained. Each candidate model was estimated from fifty random starting points. When the number of classes is relatively small, variation in BIC was relatively low from one estimation to the next. The variance in BIC rose with the number of classes. From previous analysis, it is quite likely that preferences for the contract attributes are highly correlated. Hess et al. [130] point out that latent class models incorporate this correlation inherently through class membership probabilities.

The LC-APR model used in this analysis is fundamentally different from conventional LC models in which classes are not representative of specific behaviors[123]. For this reason, Hensher et al. [125] refer to the model as a “probabilistic decision

process model” whereby the class membership probabilities represent the probability of a typical respondent exhibiting the behavior modeled in a class. Therefore, the model specification search is undertaken in a different manner. The assumptions about processing strategy are outlined first. That is, the model structure is defined. Then, the appropriate restrictions are imposed on each class defining the response strategy and then the model is estimated. In this case, the primary response behaviors of interest are non-attendance to the attributes that were left out of the market experiment and yea-saying.

For guidance on choosing likely APRs, I summarize key findings from the attitudinal section of the SP survey. Four of the five sections of the stated preference survey consisted of several questions that asked participants to rate statements about their opinions regarding their community, farms and wildlife, and farmland amenities. Most of the questions required respondents to rate the statements on a five point Likert scale from Strongly Agree to Strongly Disagree. The statements pertained to the attributes of the contracts that the participants would subsequently be comparing and were worded as follows: “Open space in agricultural use is important to me as part of Jamestown” and, “It is important to me that I can view birds and other wildlife when I walk near farms”. Nearly 97% of respondents indicated that they agreed or strongly agreed that open space in agricultural use was an important feature of their community. Undeveloped woodland was also found to be important to most of the respondents. However, the responses were mixed with regard to whether maintaining remaining agricultural landscapes was more or less important than maintaining undeveloped woodland. Participants were also asked whether they would join an expert-led bird walk if invited. There was a mix of responses to this question. A summary of these findings can be found in Table 3.2. Overall, the attitudinal findings imply that non-attendance to the Restore and Tour attributes is a distinct possibility

in the data.

The convention with regard to using latent class models with APR restrictions is to first identify candidate APRs so that they can be tested for inclusion. In many cases, the parameters are constrained to be equal across classes and attribute non-attendance is specified by restricting the parameter to equal zero in a particular class. The rationale for constraining parameters across classes is to focus on attribute non-attendance without concern for preference and scale heterogeneity. I am interested in accommodating taste heterogeneity as well as process heterogeneity. However, doing so complicates the analysis quite a bit as the combinations of potential behaviors rises exponentially. Thus, to simplify the analysis, I focus on non-attendance to attributes that were included in the SP experiment but left out of the RP treatments. While doing so facilitates tests of Hypothesis 1 above, it is also somewhat reinforced by the attitudinal findings listed above. That is, there is some qualitative evidence that individuals would not participate in a guided birdwalk if offered, and that some respondents might be ambivalent with regard to restoring fallow land to active cultivation. In addition, I let the taste heterogeneity be guided by evidence from the full-attendance classes. That is, I first tested one class for each type of APR. A layer of taste heterogeneity was added by increasing the number of full attendance classes until the lowest information criteria were obtained. I then added a layer of taste heterogeneity on to the ANA classes by assuming two of each ANA class. An outline of this process can be found in Table 3.5. The Cost ANA and All attribute ANA were confined to one class each for all specifications. The reason for not testing higher dimensions on these APRs was that constructing estimates of WTP is inconvenient for the All ANA class and practically impossible for the Cost ANA class.

For comparison against a model without APRs, I perform the classical LC model specification search by testing up to thirteen unrestricted classes. The unrestricted

model is used as a baseline for comparison of performance against a model that incorporates APRs.

Constructing Individual-level WTP estimates

In order to assess the predictive validity of the LC-APR model, it was necessary to construct individual estimates of WTP for the contract that was presented to the individuals for each year of the RP market experiment. Individual-specific conditional probabilities of class membership conditional were used to weight the within-class parameters which were then applied to the standard formula for calculation of WTP based on Hannemann's formula [135]. The procedure is outlined below.

Given the $K \times C$ class parameter estimates, I estimated the conditional probabilities of each individual belonging to each of the classes. I use these conditional probabilities as weights on the class parameters to construct the marginal utilities for each individual[132]. I then combined the estimated parameters with the RP contract attributes. I constructed choice-instance expectations of WTP by combining the individual constructed marginal utilities of acres and view, subtracting the status quo estimate, and dividing by the marginal utility of income.

I utilize within and across class variance-covariance to construct individual-specific estimates of WTP as follows:

$$\begin{aligned}
 WTP_{RP,year} &= \frac{1}{R} \sum_{r=1}^R \sum_{c=1}^C \eta_{cn}^r(\beta^r, \theta^r) \times \frac{\beta_{A,c}^r Acres_{RP,year} + \beta_{V,c}^r View_{RP,year} - \beta_{No,c}^r}{-\beta_{cost,c}} \\
 &= \frac{1}{R} \sum_{r=1}^R \sum_{c=1}^C \eta_{cn}^r(\beta^r, \theta^r) \times \widehat{SPWTP}_{RP,year,c}
 \end{aligned} \tag{3.1}$$

I generate 1,000 random draws of a multivariate normal distribution utilizing the parameters and variance-covariance matrix from the LC-APR model. For each draw, r , the individual-specific conditional probability of class membership are calculated given the parameter values, $\eta_{cn}^r(\beta^r, \theta^r)$, which are dependent upon the class member-

ship parameters θ as well as the $K \times C$ β parameter estimates. The class membership probabilities weight the expected WTP for the contract administered to the individual in a given year. The mean and 95% confidence intervals for the 1,000 draws of the individual-level values are used to analyze consistency in WTP across SP/RP treatments.

Analyzing Market Responses

I examine market participation by APR via Random Effects Probit model of the decision to return the market mailing. Building upon previous analysis [136], an individual is considered to participate in the market if she returned the market mailing, even if her offer amount was zero. I test for differences in participation rates among SP vs. RP respondents in general, and class membership probability specifically, controlling for elicitation features and other demographic characteristics. The class membership probabilities are constructed from the conditional probabilities of class membership for individuals who returned the CE survey and from unconditional probabilities of class membership based on the results of the class membership model described above and given the particular respondent's demographic characteristics.

Contingent upon market participation, I quantify differences in actual offers based on class membership, contract characteristics of the market mailing, and demographic characteristics. A Selection-Adjusted Interval Regression econometric specification is used to account for the mixture of discrete choice and payment card data. The selection adjustment is achieved via calculation of Inverse-Mills ratios (IMRs) described in Swallow et al[136] and based on Wooldridges [137] panel version of a Heckman-type selection model.

Finally, I analyze whether stated WTP are consistent with RP payments by constructing a set of rules that define "consistency" for this data. The set of criteria determine whether an individual responded consistently, or whether they under- or

over-valued the market good based on the estimates of stated WTP. The criteria depend on the type of payment format, discrete choice (DC) or open-ended (OE), the respondent received.

I define consistent behavior in this context as follows. First, if the offer amount was within the 95% confidence interval of estimated willingness to pay from the choice experiment (SPWTP) for individual i , then she has made an offer that is consistent with her choices in the CE survey. There is one other case in which I consider behavior to be consistent. That is, if the lower bound of the 95% confidence interval on \widehat{SPWTP} is above the highest dollar value in the range of the open-ended treatment or the point value in the discrete choice treatment, and that value is chosen as the offer amount, I consider this behavior to be consistent.

Inconsistent choices are of two types. If behavior is not found to satisfy the two previous criteria, then the SP estimates overvalue the market contract if the lower bound of the 95% CI is greater than the offer amount. I identify these individuals as exhibiting hypothetical bias. Alternately, if the upper bound of the 95% CI is below the offer amount, the respondent is considered to have under-valued the good in the SP scenarios.

3.4 Results

Descriptive Statistics and Class Allocation

Upon preliminary examination of the SP experiment data itself, there appeared a few clear strategies. For instance, the yea-sayers (those who answered that they would purchase both contracts for all questions) comprised a substantial share of the overall respondents (13.8%). There was also evidence that many individuals were simply choosing the lowest cost option (10.7%). These qualitative findings were used to examine the performance of the LC-APR model.

The final LC-APR model was arrived at after methodical testing. The best model in the specification search, based first on BIC and then on highest average maximum conditional class membership probability had five classes: four restricted classes and one full attendance class (Table 3.6). The final five-class LC-APR model included the following APRs: Restore and Tour ANA, Tour ANA, All ANA, Cost ANA, and one full attendance (unrestricted) class.

The LC-APR model results suggest that full attendance was not a majority strategy (Table 3.7). Individuals are assigned to classes based on highest conditional probability of membership. There were 173 respondents (22%) for whom full attribute preservation was a best fit. All attributes in the full attendance, Restore/Tour attribute non-attendance, and Tour non-attendance classes are significant and have the expected signs. The All attribute non-attendance class and Cost ANA class parameters are insignificant. Individuals who fit these classes with high conditional probability act in a way that makes it difficult for preference trade-offs to be calculated with accuracy.

Individuals who chose both contracts for all choice experiment questions (the yeayers) fit the Cost ANA class with extremely high conditional probability (at least 90%). The remaining seven individuals in the Cost ANA class made only one choice that was not the Both alternative. The unfortunate drawback of this class is that constructing willingness to pay estimates involves division by the cost parameter which is restricted to zero. Therefore, I found that the estimates of WTP for these individuals, using the method described in Equation 3.1, produced wildly high negative or positive values because the cost parameter at the individual level was exceeding close to and equally likely to fall on either side of zero. In addition, confidence intervals around these estimates ranged in the thousands.

Following some authors, I included a class that represented full attribute Non-Attendance (All ANA) to capture idiosyncratic behavior. There were 82 respondents who fit this classification best. A portion of these respondents were found to have extreme reactions to a change from the referendum vote to one of the elicitation mechanisms.¹ Fourteen of the respondents (17%) were “protest votes”, that is, they chose the No Buy option for all choice occasions. The remainder of the All ANA class exhibited seemingly random response behavior. The parameter estimate for the Cost parameter for this class is positive and very close to zero, leading the average WTP to be negative and drastically large.

There were 420 individuals who fit the Restore/Tour ANA and Tour ANA classes best. Recall that the primary goal of modeling a Restore/Tour ANA class is to test whether individuals who were relegated to this class made choices that were more consistent in the market experiment, which did not include these contract features. This class had a high positive value on the View attribute relative to the other classes, perhaps implying lexicography on this contract characteristic and indicated that a segment of the population is highly concerned with preserving the aesthetics of unharvested farmland. An alternative interpretation is that, if a parcel can be viewed from the road, then ensuring that farmers uphold the contract is possible. Monitoring compliance on a parcel that cannot be seen from a road would be difficult. In the membership model of the LC-APR, membership to this class was comprised of individuals who had high values of the Equality variable (Table 3.8). This variable was constructed from responses to the analysis of the survey responses regarding preferences for fair and equal payments for ecosystem services.

¹The SP experiment sample was split into two main groups. Group 1 was issued CE treatments in which the respondents made choices under two separate scenarios: a referendum vote and the prescribed elicitation mechanism. Group 2 was not administered the referendum and thus made all five choices given a single elicitation mechanism

The second class (Tour ANA only) had low but positive values for the attended attributes. Participants who consistently chose the cheaper alternative fit this class best. Ideally, the LC-APR model would have relegated lexicographic preferences on cost to an All ANA class. This fact lends support to a further refinement in which an indicator variable for the lower cost alternative and lexicography on this variable might be included as a class. The refinement was not implemented here because it would lead to a rise in the number of individuals for which estimates of WTP would not be possible. In essence, it is likely not accurate to assume that individuals who fit this class with high probability were in fact ignoring the Tour attribute but that instead these individuals have overall low values for the farm-wildlife contract features. A summary of the classes with average WTP for an average sized parcel with a view is listed in Table 3.7.

For comparison, an unrestricted latent class model was estimated (Table 3.9). The six class unrestricted model succeeded in relegating the yea-sayers to a single class (Class 3) as well. This feature highlights the advantage of the LC model as a non-parametric representation of taste heterogeneity. If conventional mixed logit estimation was performed, then the commonly exercised assumption of normally distributed coefficients means that the population-level estimates of the parameters would be pulled up by the influence of these responses.

Market Experiment Outcomes

Participation

The Random Effects Probit model of market participation (Table 3.10) measured the effects of demographics, contract characteristics, and first-stage class membership on the decision to return the market experiment survey. An individual is considered to have participated in the market experiment if she returned the payment card, even if the amount offered was zero. Consistent with previous analysis, significant demo-

graphic determinants of participation were the respondent's age, donation history, and a history of mail-order from children's catalogues. Age and a history of donation both increase the likelihood of returning the market solicitation. Mail-ordering from children's catalogues decreases the likelihood of participation, presumably because household resources of time and money are more limited for this group.

Of the contract attributes found to influence participation, only the log of the minimum amount of the solicitation was found to be significant. As in Swallow et al[136], I acknowledge that this implies that at least some of the individuals did open the solicitation and look at it before deciding not to return the mailing. None of the effects-coded variables for mechanism treatment were found to contribute to the decision to participate in the market experiment.

In order to investigate participation given the first stage SP results, I included the conditional probability of class membership for each of the respondents estimated from the LC-APR model and the unconditional probabilities of class membership for the market experiment respondents who did not return a CE survey. These measures capture the probability of membership into an APR class, conditional on CE choices if available and unconditional probabilities for respondents who returned the market experiment only. Again, the unconditional probabilities are influenced by the LC class membership model which takes account of demographic attributes (Table 3.8). Consistent with the qualitative findings (Table 3.4), having returned the choice experiment survey significantly increased the likelihood of returning the market mailing. The full attendance class was omitted for identification purposes. Negative and significant parameter estimates on the class membership variables implies that individuals who were likely to attend to all attributes in the SP survey had the highest likelihood of participation in the market experiment relative to the other classes. Interestingly, high probability of membership in the third class (All ANA) yields the

lowest participation rate. This class included all individuals who chose the No Buy option for all questions and so this finding is fairly consistent with expectations given the LC results. Individuals who belong with high probability to the Cost ANA class were not found to have a participation level that was significantly different than the full attendance class. As these individuals mostly chose the highest level of contribution in the SP survey, this result is not surprising.

Offers

Contingent upon participation in the market, the selection-adjusted interval regression of offer amount results indicate that class membership and contract payment format had important impacts on the amount of money offered in payment toward the farm-wildlife contracts (Table 3.11). Individuals who received a discrete choice (DC) solicitation on average made much higher offers than those who were issued low range open-ended (OE) payment cards (\$10 - \$80). Those who were issued the higher range payment cards (\$35 - \$120) contributed at a level intermediate to the low OE and DC treatments. This finding, robust to alternate specifications, is consistent with findings from previous studies regarding differences between payment card and discrete choice elicitation mechanisms. Other aspects of the contracts themselves were not found to significantly impact offer amounts.

Most of the class membership regression coefficients have significant explanatory power. The parameter estimates are to be interpreted as differences from the set of individuals who fit the full attendance class with high likelihood. Relative to full attendance, all LC-APR classes are found to offer lower amounts in the market experiment. According to these results, the order of offers from highest to lowest was

$$FullAttend > CostANA > TourANA > AllANA > Restore/TourANA \quad (3.2)$$

compared to what the LC-APR predicts

$$CostANA > FullAttend > TourANA > Restore/TourANA > AllANA \quad (3.3)$$

Interestingly, the respondents who fit the Cost ANA class best (yea-sayers) were found to make lower offers than individuals who were likely to attend to all attributes. Because the cost range was nearly identical across CE and market experiment treatments, the level of hypothetical bias in this group is evident. That is, if yea-sayers were to behave in a manner consistent with their responses in the hypothetical survey, one would expect these individuals to choose the highest possible offer amounts in revealed preference situations. In fact, the results show that, while the average offer is high for this group, it is not the highest. Thus, if full attendance implies that these individuals invested more in terms of cognitive effort, this investment manifests as higher WTP in the market.

The Restore and Tour ANA individuals revealed lowest payments in the market experiment but also lowest positive values in the choice experiment. Therefore, this result is not particularly surprising. Similarly, members of the Tour ANA class had low stated values for all attended attributes, but, were not found to consistently choose the No Buy option. Thus, it follows that these individuals also exhibit low revealed values.

Measuring Hypothetical Bias and SP undervaluation

Based on the rules of consistency outlined above, I compare the LC-APR model against the LC model with preference heterogeneity alone (Tables 3.12 and 3.13). The average value of the hypothetical bias for the LC attribute processing model was \$37.69 compared to \$52.79 for the unrestricted model. This points to clear evidence that the LC APR model performs better at mitigating the effects of hypothetical bias (Hypothesis 3). Both models produced similar results with regard to undervaluing

the market good. Of the respondents who were found to undervalue the market good in the SP treatment, 3 chose the lowest value on the payment card and about half were issued a discrete choice solicitation. This may exemplify the advantage that the DC format maintains over the open-ended format: that some individuals may feel pressured to participate and will accept a discrete choice payment out of this compulsion.

Focusing on the results of the attribute processing model gives a sense of possible sources of hypothetical bias. The highest proportion of consistent values can be found for the class of individuals who fit the Restore and Tour ANA class best. This supports the first hypothesis of this research in finding that individuals who are found to be rather insensitive to the attributes that do not transfer to the market can be expected to produce more consistent or reliable estimates of willingness to pay. In addition, the Cost ANA class contained the highest proportion of respondents who exhibited hypothetical bias. Because calculating a realistic estimate of WTP for this class was impossible, consistency for the Cost non-attendance class was defined as choosing the highest level on the open-ended treatment, accepting the discrete choice price, or writing in an amount above the range in both cases. Under this definition of consistency, 67% of yea-sayers exhibited hypothetical bias. Of the consistent responses for this group, each individual chose the highest value presented them. This is 1.5 times the amount of hypothetical bias exhibited by the full attendance class which had the next highest proportion. Therefore, I find some support for Hypothesis 2, that yea-sayers reveal a high level of hypothetical bias but also exhibit high valuation otherwise.

3.5 Summary and Conclusions

Stated choice experiments are good tools for assessing the nature of potential demand for a new product. The strength of the discrete choice experiment lies in the

ability of the researcher to generate data that contains sufficient variation so as to best infer the nature of the trade-offs among attributes that potential market participants will make. However, unlike new product development in the realm of tradable goods, a significant challenge to the success of eliciting preferences for nonmarket goods is inducing respondents to reveal their true values for a nonexclusive public good.

Behavioral economists and choice modelling researchers have found mounting evidence that the neoclassical model of utility maximization does not always succeed in characterizing economic behavior. Attribute processing strategies have an effect on valuation in CE surveys and that these processing strategies have implications for market participation and contribution in this unique dataset. It has been suggested that accounting for such strategies may help alleviate some of the well-documented differences in measured stated vs. revealed values, especially for public goods.

The study outlined in this paper offered a unique means by which to examine differences between revealed and stated preferences for two reasons. First, the CE survey included an option to purchase both contracts, thereby releasing the respondent from the need to make any trade-offs. Individuals who were inclined to choose this option were found to be of two types: respondents with high values for the contracts and their attributes, and respondents who exhibited hypothetical bias. By comparing the behavior of these respondents in the subsequent market experiment, I was able to clearly observe the level of HB resultant from this strategy of hypothetical response. My findings indicate that yea-sayers may be expected to exhibit high contribution levels, but overall, were not found to participate to a larger extent than other respondents. This implies that some means of partitioning this group into individuals whose actual value for the good is high and those who are engaging in yea-saying would assist in determining how much weight to assign these responses when making inferences about public goods values.

In addition, the CE survey involved the presentation to respondents of a ‘good’ that was not previously available to the community. Rather than testing a new dimension of an already existing consumption good (i.e. an added label to an existing carton of milk or a new transportation option), the survey involved assigning value for a new product with possibly unfamiliar characteristics. Therefore, applying simplifying rules to make the choice easier was a distinct possibility. Indeed, I find that a model that incorporates APRs succeeds in mitigating the upward bias of SP responses.

The method employed in this research can be useful for many applications. In general, future research may utilize these findings to test for shortcomings of the survey design or in identifying and calibrating for specific processing heuristics so that a more accurate assessment of values can be derived from hypothetical survey data. With the growing emphasis on market-based solutions for ecosystem services provisions, it is increasingly important to accurately measure values and identify challenges at the market design stage. This research provides some preliminary evidence for what works with regard to getting the prices right.

Table 3.1. Summary of Choice Experiment contract attributes

Attribute	Description	Levels
Acres	Number of acres to be placed under contract upon which farmer will delay mowing and harvesting	10, 25, 40, 55
Restore	Number of acres to be restored to active hay fields, not restricted to delayed mowing/harvesting	0, 10, 20, 30
High Bobolink	Level of expected fledglings saved (correlated with acreage)	Low, High
View	Whether the parcel will be viewable from the road	View, No View
Tour	Whether individuals paying into the contract are invited to an expert-led birdwalk	Tour, No tour
Cost	The cost of the farm-wildlife contract	\$10 — \$105

Note: There were five defining attributes of each of the CE contracts. The attributes and their levels are listed above. Source: [76].

Table 3.2. Summary of Attitudinal Findings from Supplementary CE Questions

Question	Summary of Response
<i>Open space in agricultural use is important to me as part of Jamestown</i>	97% Agree or Strongly Agree
<i>Undeveloped woodland is important to me as part of Jamestown</i>	Over 90% Agree or Strongly Agree
<i>It is important for me to maintain remaining agricultural landscapes than to maintain undeveloped woodlands in Jamestown</i>	Results indicate that there is some heterogeneity with respect to this issue
<i>It is important to me that I can view birds and other wildlife when I walk near farms</i>	Highly skewed toward Agree or Strongly Agree
<i>Improving farm management to provide Bobolink habitat</i>	Majority response was “somewhat valuable” with significant heterogeneity
<i>Improving public access and educational opportunities such as walking tours to visit farms and watch wildlife</i>	Over 79% of responses in somewhat to extremely valuable, with the rest in the somewhat to not at all

Note: The results from a selection of the attitudinal section of the Hypothetical Choice experiment are listed above. Source: [77]

Table 3.3. Comparison of Attributes and Levels across SP and RP treatments

Attribute	Choice Experiment	Market 2007	Market 2008
Field size to delay harvesting	10 – 55	10 – 18	10
Acres of farmland restored to cultivation	0 – 30		
Bobolink Concentration	Low, High		
# of 2006 territories		1 – 4	
# of Fledglings			6—10, 10—14, 14—18
View from road	View, No view	View, Partial, No view	View, No view
Mechanism	SP, VCM:PPMBG, PR, PM, UPA	PM, PR, UPA	PM, PR, UPC
Cost	\$10 — \$105	\$10 — \$120	\$10 — \$120

Note: The mechanisms were as follows: SP: referendum for tax increase, VCM:PPMBG: provision point mechanism with money-back guarantee, PR: provision point with money-guarantee and proportional rebate of excess funds, PM: pivotal mechanism, UPA: uniform price auction, UPC: uniform price cap.

Table 3.4. Summary of Market Response

	Total		SP respondents		SP Yeah-sayers		RP only	
	2007	2008	2007	2008	2007	2008	2007	2008
N	2791	2680	764	713	105	101	2027	1967
Returned Market mailing	13%	8%	30%	16%	34%	14%	7%	5%
Made Offer	7%	5%	19%	11%	27%	11%	3%	3%
Average Value of Offer	\$47.94	\$46.49	\$50.30	\$46.98	\$60.54	\$41.82	\$42.11	\$45.83

Note: The table is divided into four groups: all individuals who returned the market mailing, those who returned both the market and CE surveys, the subset of CE respondents who chose the “Both” option for all questions, and all other individuals who were mailed a market experiment survey.

Table 3.5. Specification Search

Non-Attendance	Number of Classes								
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Restore/Tour	1	2	1	1	1	2	1	1	2
Tour	1	1	2	1	1	1	2	1	2
All	1	1	1	2	1	1	1	2	2
Cost	1	1	1	1	1	1	1	1	1
None	1	1	1	1	2	2	2	2	2
BIC	7137	7149	7149	7151	7142	7172	7176	7166	7249
AIC	6904	6863	6859	6875	6847	6826	6825	6829	6805

The specification search is outlined above. Each column represents the number of classes specified in each model to have the parameter restrictions that imply the non-attendance patterns listed in the first column. Boldfaced values represent the best performing information criterion.

Table 3.6. Latent Class APR Model Results: Five Classes

	Class 1	Class 2	Class 3	Class 4	Class 5
	Restore/Tour	Tour	All	Cost	Full Attend
Cost	-0.074***	-0.045***	0.001	0	-0.017***
No ASC	-0.487***	-1.582***	-1.168	-8.718	-0.871***
Acres	0.011**	0.028***	0	0.106	0.02***
Restore	0	0.026***	0	0.13	0.055***
View	0.305***	0.124**	0	8.95	0.211***
Tour	0	0	0	0.095	0.3***
Share	0.159	0.386	0.085	0.145	0.225
BIC			7137		
AIC			6902		
LL			-3401		

Notes:*** p<0.01, ** p<0.05, * p<0.1

Table 3.7. Summary of Classes

Class	Definition	Unique Features	Avg. WTP for 10 acres with view	N
1	Restore & Tour ANA	Possible Lexicography on View	\$12.19	120
2	Tour ANA	Includes individuals who chose cheaper contract consistently	\$44.13	300
3	All ANA	Protest Votes, Extreme reactions to payment mechanism, quixotic behaviors	\$-1168.00	82
4	Cost ANA	Includes all yea-sayers	NA	116
5	Full Attendance	High Values	\$75.41	173

Note: N is the number of individuals for whom the maximum conditional probability places them in the class

Table 3.8. Latent Class APR Membership Model

Variable	Class 1 Restore/Tour	Class 2 Tour	Class 3 All	Class 4 Cost
Age	0.01	0.01	0.018	0.004
Income	-0.004	-0.001	-0.000	-0.002
Equality	0.929***	0.186	-0.130	-0.290
Mail Order Kids	-0.256	0.05	-2.42	-0.170
Education	-0.549***	-0.311***	-0.934	-0.112
Gender	-0.397	-0.05	-1.07	-0.353
Constant	2.45**	1.79***	3.50	0.251

Notes:*** p<0.01, ** p<0.05, * p<0.1

Class 5 is omitted from estimation for purposes of identification.

Table 3.9. Latent Class, Unrestricted Model

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Cost	-0.049***	-0.016***	-0.126***	-0.094***	-0.015***	-0.007
No ASC	-1.537***	0.107	-7.783	-0.651***	-1.057***	-2.175***
Acres	0.033***	-0.032***	1.869***	0.01***	0.037***	0.008
Restore	0.024***	0.024*	-0.187***	0.034***	0.075***	0.03***
View	0.1*	0.96***	-0.609	0.297***	0.221**	0.124
Tour	0.094*	0.072	7.531	-0.224**	0.241**	0.279***
Share	0.36*	0.038***	0.131	0.015	0.156	0.166
BIC				7096		
AIC				6904		
LL				-3411		

Notes:*** p<0.01, ** p<0.05, * p<0.1

Table 3.10. Participation Equation

	Coeff	z
Constant	-2.881***	-5.380
Contract Attributes		
Acres	0.00423	0.261
View	0.111	1.086
# Bobolink Territories	0.0137	0.279
# Fledglings	0.003	0.120
lnMinAmt	-0.0982*	-1.849
Discrete Choice	0.0874	1.133
Mechanisms		
Uniform Price Auction	0.0639	0.697
Pivotal Mechanism	0.0342	0.530
Uniform Price Cap	-0.128	-1.123
Class Membership		
Returned SP	1.212***	11.54
$P(\text{RestoreTourANA} CE)$	-0.454*	-1.864
$P(\text{TourANA} CE)$	-0.369*	-1.700
$P(\text{AllANA} CE)$	-0.987**	-2.442
$P(\text{CostANA} CE)$	-0.152	-0.619
$P(\text{FullAttend} CE)$	-	-
Other Demographics		
Log Purchasing Power	-0.0982*	-1.849
Age	0.0167***	4.184
Donation History	0.260***	2.585
Environmental Donations	0.135	0.976
Mail Order Kids Mags	-0.304***	-2.782
Year 2008	-0.459	-1.639
$\ln\sigma_u^2$	0.366**	2.191
LL	-1579	
N	5415	
χ^2	180.38	
p	0.000	

*** p<0.01, ** p<0.05, * p<0.1

Note: The class membership variables represent the conditional probability of class membership given the individual's responses to the choice experiment survey.

Table 3.11. Payment Equation

	Coeff	z
Constant	21.99	0.269
Contract Attributes		
Discrete Choice	39.62***	5.197
OE High	18.37**	2.471
Acres	0.0351	0.0327
View	8.249	1.212
# Bobolink Territories	-3.511	-1.089
# Fledglings	2.504	1.443
Mechanisms		
UPAe	-9.901	-1.601
PMe	4.541	0.988
UPCe	-1.399	-0.180
Class Membership		
Returned SP	22.58***	3.773
$P(\text{RestoreTourANA} CE)$	-91.21***	-5.786
$P(\text{TourANA} CE)$	-45.64***	-3.363
$P(\text{AllANA} CE)$	-60.67**	-2.369
$P(\text{CostANA} CE)$	-26.97**	-1.991
$P(\text{FullAttend} CE)$	–	–
Other Demographics		
LogPurchasing Power	4.18	0.775
Age	0.0225	0.0603
Donation History	-1.605	-0.209
Environmental Donations	7.825	0.833
Missing Donation Information	22.47	1.154
Mail Order Kids Mags	-31.29***	-3.849
Inverse Mills and 2008 Indicator		
2008	-36.58	-0.775
IMR	4.518	0.129
IMR 2008	3.699	0.126
σ_u	46.36***	12.10
σ_e	23.60***	5.937
LL	-740.63	
BIC	1652.88	
N	576	
χ^2	94.11	
p	0.000	

*** p<0.01, ** p<0.05, * p<0.1

Note: The class membership variables represent the conditional probability of class membership given the individual's responses to the choice experiment survey.

Table 3.12. RP Offer Evaluation - LC APR Model

	Consistent Values	Hypothetical Bias	Under Valued	Total
Restore/Tour ANA	31	3	12	46
Tour ANA	52	49	28	129
All ANA	13	4	1	18
Cost ANA	20	35	0	55
Full Attendance	31	36	17	84
Total	147	124	58	332

Note: Individuals are relegated to a class based on the highest conditional probability of class membership. Consistent responses indicate that the individual's offer lies within the confidence interval of their estimated WTP from the choice experiment. Hypothetical bias exists when the lower bound of an individual's estimated WTP lies above the market offer amount. Similarly, under valuing occurs when individuals state a lower WTP in the choice experiment than offered in the market experiment.

Table 3.13. RP Offer Evaluation - Unrestricted Model

	Consistent Values	Hypothetical Bias	Under Valued	Total
Class 1	50	48	31	129
Class 2	2	5	2	9
Class 3	20	31	0	51
Class 4	32	3	11	46
Class 5	27	23	4	54
Class 6	13	21	9	43
Total	144	131	57	332

Conclusion

The overarching goal for this research was to examine how market mechanisms can be used in integrating values for ecosystem services into the economy. As consumers and policy-makers become more and more concerned about deleterious impacts of human activities on the environment and the benefits that the environment bestows, increasingly sophisticated means of measuring these values are being explored by both public and private institutions. Market mechanisms for private goods and services can be relied upon to efficiently allocate goods and services among societies members. However, there are many issues involved in assessing the values of environmental goods and services that are not present in markets for private goods.

I examined these issues through both stated and revealed preference methods. In all three manuscripts, I focused on demand-side considerations: that is, measuring and evaluating the willingness-to-pay for environmental goods and services. There are several important considerations on the demand-side with regard to: identifying environmental goals that resonate with consumers, overcoming the tendency to free-ride on others contributions, and over-stating one's value for environmental goods in hypothetical settings. Each manuscript highlights a different aspect of these features of environmental valuation.

In manuscript 1, I measured the effectiveness of eco-labels to generate value for products made from fish that was certified sustainably harvested by the largest certification agency for fisheries, the Marine Stewardship Council (MSC). I used scanner data from the London metropolitan area of the United Kingdom. I chose to focus my analysis on certified Alaskan pollock products because the Alaskan pollock fishery was one of the first fisheries certified by the MSC and had a significant presence in supermarkets in the United Kingdom. The mere existence of the eco-label on shelves

in supermarkets does not necessarily guarantee that market benefits are being generated as a result. In fact, in comparison to other labels that consumers may conceive to convey information about health benefits of the product, such as organic labeling, interest in promoting sustainable fishing might not motivate consumers to shift consumption practices toward the labeled product. While there was evidence from stated preference experiments that individuals value the label and the cause, this was the first study to quantify the realized market benefit that the label generates. I found evidence of a 14.2% price premium for products that contain MSC-certified Alaskan pollock in this market.

In comparison to other studies that assess preferences for the MSC's eco-label within the United Kingdom, these results are similar in magnitude to those found for salmon products, 13.1% [64] and chilled haddock products, 10% [63]. However, as Johnston et al. [5] have pointed out, there are significant cross-cultural differences in preferences. Therefore, a comparison of our findings with findings from other markets for pollock products would produce a more comprehensive assessment of the market benefits of certification. The closest comparison in the literature is the finding of Uchida et al. [138] that consumers in Japan are willing to pay a 20% premium for certified salmon products if given information about the label. These findings may suggest that more work is needed in bringing issues of sustainability to consumer consciousness, at least for some markets. A comprehensive assessment that includes the main markets for Alaskan pollock products would yield a more accurate estimate of the price premium that the MSC's label on pollock generates.

In the second manuscript, I examined a stated preference survey administered to the inhabitants of Jamestown, Rhode Island, designed to elicit values for the preservation of grassland nesting habitat on Jamestown's farmland. The Jamestown experiment was unique in that several different public goods auction mechanisms were tested in

the field. Because the environmental good had not been previously offered to the Jamestown residents, I was interested in ways of capturing uncertainty in response to the survey and how the elicitation mechanisms, which were designed to mitigate the urge to free-ride, might complicate decisions in field studies for public goods. Interestingly, while I did not find evidence that more esoteric elicitation mechanisms added noise to the participants responses, I found some evidence that adding a familiar base mechanism prompted respondents to behave in a manner that was more consistent with findings in the laboratory with regard to these mechanisms.

The performance of voluntary contribution against provision point mechanisms has been tested in the laboratory and in the field[139, 10]. As Rose et al [10] point out, provision point mechanisms provide product definition which has been judged by market researchers to assist participants in understanding what they will receive in return for their offers. While all of the mechanisms in the Jamestown choice experiment were administered with both provision points and money back guarantees, this was the first time to my knowledge that the pivotal mechanism which has been proven theoretically incentive compatible was tested in the field. I found encouraging evidence that, contrary to many assertions, demand-revealing mechanisms can be used in field settings to raise contributions and reduce free-riding. But there is significant evidence to warn future researchers that individuals will ignore the elicitation mechanisms if not emphasized. It is unclear whether the placement of the mechanism in the middle of the choice task or the application of a familiar reference mechanism was what induced respondents to have stronger reactions to the mechanisms but it is clear that emphasis must be placed on the novelty of the mechanism if it is to perform as expected in the lab, something that is akin to the researcher reading instructions aloud in the lab.

Stated preference surveys are often criticized for their tendency to suffer from hypothetical bias. Recent research has voiced concerns about the usefulness of hypothetical choices at all given this well-documented drawback [1]. Carson and Groves' seminal paper [140] addressed hypothetical bias in relation to the perception of "consequentiality" of an individual's response. According to Vossler and colleagues [141], "a survey elicitation is consequential if the respondent cares about the policies contemplated ... and further views her response as potentially influencing agency action". It follows logically that respondents may evaluate the elicitation mechanisms as influencing the consequentiality of their responses. This seems to be built in to the structure of the pivotal mechanism. If it is the case that individuals view their responses to be more or less consequential as a result of the mechanism administered, then a variety of levels of hypothetical bias can be expected for this data. This may have a confounding effect on our results. In fact, Swallow et al.[136] found qualitative evidence that offers were higher in uniform price cap and proportional rebate than for pivotal mechanism in the revealed preference experiments of the Jamestown project. The failure of the pivotal mechanism to generate higher offers may indicate some distrust in the mechanism in practice. More work is needed to investigate these dynamics.

In the final manuscript, I utilized the findings from the stated preference survey and compared them against the subsequent two-year market experiment that was administered to the same community. Care was taken to match elicitation mechanisms as best as possible in both phases of the project. I hypothesized that respondents would be more inclined to use simplifying rules such as yea-saying and attribute non-attendance when answering hypothetical questions regarding the market. My interest was in the implications of these findings for behavior in the revealed preference market experiment. I was able to identify several processing strategies and, as expected, a

higher rate of hypothetical bias in individuals who chose the highest level of provision in all choice occasions in the stated preference survey. I have shown that attribute non-attendance can be used to identify sources of hypothetical bias as a result of survey design and respondent type. I have also found that the methodology can be used to identify estimates that will yield more reliable results. This was the first study that took the results of burgeoning research on attribute non-attendance and tested their implications for realized behavior.

These findings have implications for survey and market design particularly for public goods valuation. Researchers are becoming increasingly concerned with identifying the incentive structure of stated preference methods when formulating expectations pertaining to hypothetical bias [140, 2]. In many ways, the work done by Carson and Groves is reshaping our thoughts with regard to the issue. There may be a link between the attention paid to the survey and the degree of consequentiality perceived by the respondent. Cameron and DeShazo [25] and others point out that individuals may attend to some attributes more or less than others as a result of the cognitive effort of the choice task. If a respondent is found to attend to all attributes, then it can be rationally assumed that she cares more about the initiative and thus deems her responses as consequential. Indeed, I found that individuals who were likely to fully attend to all choice experiment attributes had higher revealed offers in the market experiment. However, I found only weak evidence that participants who fully attended to all attributes of the choice experiment exhibited less hypothetical bias in the revealed preference scenario. Instead, accounting for non-attendance to attributes that did not transfer to the revealed preference domain significantly reduced measured hypothetical bias. That is, our results suggest that there may be more gains to be had in terms of consistent estimates of WTP by focusing on respondents who care very little for the attributes that are not transferable to the revealed preference context.

Regardless, the results of this manuscript suggest promising new directions for the use of attribute non-attendance research in survey and new market design. Research on attribute processing strategies is likely to prove invaluable for survey design and analysis

There were some notable limitations to this research and I thank several conference attendees and committee members for their input regarding these issues. First, generating aggregate estimates of marginal willingness to pay using this method would require sample selection adjustment at multiple levels, first to correct for self-selection in response to the stated preference survey and then for each year of the market experiments. While we attempted several variations on corrections for sample selection bias, none were found to be satisfactory. Further research is needed to remedy this drawback. In addition, while I identify individuals who employ different processing strategies, some way of accounting for this information at the level of the population would enhance valuation estimates.

Overall, the research presented in this dissertation provides some insights into how effectively markets for private goods are being used to provide environmental goods and services and how new, direct market mechanisms for providing these amenities can be employed in the field. Markets have the potential to provide an important complement to government programs to enhance ecosystem services if efficient mechanisms to reduce free-riding behavior can be developed. Further research is needed to advance our understanding of how public goods elicitation mechanisms can be successfully transferred from the lab to the field and, if not, how adjustments can be made.

APPENDIX A

Elicitation Mechanisms Employed to Mitigate Free-Riding

This section provides a verbatim transcription of the public goods elicitation mechanisms administered in both the stated and revealed preference experiments.

A.1 The Stated Preference Mechanism Descriptions

Figure A.1 provides an example of the layout of the mechanism descriptions and describes the voluntary contribution mechanism. The other mechanisms administered in the stated preference survey are subsequently listed verbatim.

Figure A.1. The Voluntary Contribution Mechanism

Instructions for Part 2—Please read carefully.

We will ask you to tell us **whether you would buy a share of a farm wildlife contract for the proposed amount**. Each farm wildlife contract is distinguished by the five characteristics described in the previous table.

If you agree to pay the proposed amount, you pay *only* if the total amount committed by all the Jamestown residents is **enough to cover the total contract cost**.

Our guarantee: If the total amount committed by all the Jamestown residents is **not enough to cover the total contract cost**, then we would *not* establish a farm wildlife contract and the farm business would be unable to change their management plans to protect grassland nesting birds like the bobolink. *You would pay nothing even if you had offered to buy a share.*

Q&A on how this method works

To illustrate, we use *unrealistic* numbers. Suppose we ask you whether you are willing to buy a contract-share by paying **at most \$5**.

Q1. What happens if I offered to buy a contract-share for \$5 and the total amount committed by residents **is enough** to cover the total contract cost?

Answer: You pay \$5 and we would establish a farm wildlife contract.

Q2. What happens if I offered to buy a contract-share for \$5 but the total amount committed by residents **is not enough** to cover the total contract cost?

Answer: We would *not* enter into a wildlife management contract with the farm. You would pay nothing even if you had offered to buy a contract-share.

Your task—important!

- We will ask you to compare six sets of different farm wildlife contracts.
- Please make your decision on each comparison separately, as if they were the only comparison you had today.

The other other mechanisms were worded as follows:

The Provision Point Mechanism with Proportional Rebate

We will ask you to tell us whether you would buy a share of a farm wildlife contract for the proposed amount. Each farm wildlife contract is distinguished by the five characteristics described in the previous table. If you agree to buy a share, you pay *only* if the total amount committed by all the Jamestown residents is enough to cover the total contract cost.

Our guarantees:

1) If the total amount committed by all the Jamestown residents is not enough to cover the total contract cost, then we would not establish a farm wildlife contract and the farm business would be unable to change their management plans to protect grassland nesting birds like the bobolink. *You would pay nothing even if you had offered to buy a share.*

2) If the total amount committed by all the Jamestown residents exceeds the total contract cost, then **we would give a rebate to those who offered to pay**. The rebate would be in proportion to the amount each person committed.

Q& A on how this method works

To illustrate an example, we use *unrealistic* numbers. Suppose we ask you if you are willing to buy a contract-share by paying **at most \$5**.

Q1. What happens if the amount committed by all the Jamestown residents is **more than enough to cover the total cost** of a farm wildlife contract?

Answer: Suppose you offered to buy a share for \$5 and enough other residents bought shares so that we collected 30% more than needed for the contract. Then we would give you a rebate of \$1.50 from your \$5 offer, and we would establish a farm wildlife contract.

Q2. What happens if the amount committed by all the residents is **not enough to cover** the total contract cost?

Answer: We would not enter into a wildlife management contract with the farm. You would pay nothing even if you had offered to buy a contract-share.

The Uniform Price Auction

We will ask you to tell us whether you would buy a share of a farm wildlife contract for the proposed amount. Each farm wildlife contract is distinguished by the five characteristics described in the previous table. The amount you commit would be a cap (the maximum) on how much you would pay.

The amount you *actually* pay (the “price”) would be determined after we know how much everyone else offers. The price would be the lowest dollar amount that we can identify among the residents such that we collect enough funds to pay for the farm wildlife contract **if everyone who**

agreed to pay at least that amount would pay the same price. We would then bill you this amount.

However, **if not enough residents offer to buy a share, we may not be able to identify such a price.** In that case, we would not establish a wildlife management contract with the farm, the farm business would be unable to change their management plans to protect grassland nesting birds like the bobolink.

Our guarantee: If we cannot identify a “price” as described above, we would not establish a wildlife management contract. In that case, *you would pay nothing even if you had committed to buy a share.*

Q&A on how this method works

To illustrate an example, we use *unrealistic* numbers. Suppose we ask you if you are willing to buy a contract-share by paying **at most \$5.**

Q1. What happens if I commit **\$5** and the “**price**” (the lowest amount someone is willing to pay that would meet the total contract cost if everyone who buys would pay this amount) is **\$2**?

Answer: You get a rebate of \$3 and we would bill you just \$2. Everyone who agreed to buy a share for at least \$2 would pay \$2 and we would establish a farm wildlife contract.

Q2. What happens if I commit \$5 but the “**price**” could not be identified?

Answer: If not enough residents offer to buy a share at a large enough cap, then a “price” may not be identified. In this case, we would not establish a farm wildlife contract. You would pay nothing even if you had offered to buy a contract-share.

The Pivotal Mechanism

We will ask you to tell us whether you would buy a share of a farm wildlife contract for the proposed amount. Each farm wildlife contract is distinguished by the five characteristics described in the previous table.

What you *actually* pay depends not only on your decision but on the **total amount committed by all the other Jamestown residents.**

If you offer to pay the proposed amount, and:

- **if your payment is required** to meet the total contract cost after we have added up all other residents’ offers, then we would bill you for the entire payment. We would establish a farm wildlife contract.
- **if your payment is not required** to meet the total cost, i.e., the total of offers from all other residents is enough to meet the total contract cost, then *you pay nothing even if you had committed to buy a share.* We would establish a farm wildlife contract.

Therefore, **you pay only if your payment makes the difference between meeting the total cost and not meeting the total cost.**

Our guarantee: If the total amount committed by all the Jamestown residents is less than the total contract cost, we would *not* enter into a wildlife management contract and the farm business would be *unable* to change their management plans to protect grassland nesting birds like the bobolink. *You would pay nothing even if you had offered to buy a share.*
Q& A on how this method works

To illustrate an example, we use *unrealistic* numbers. Suppose we ask you if you are willing to buy a contract-share by paying **at most \$5**.

Q1. What happens if the total of offers from all other residents fall short by an amount less than or equal to my payment (\$5)?

Answer: You would pay \$5. We would enter into a farm wildlife contract.

Q2. What happens if the the total of offers from all residents (including me) is not enough to meet the total contract cost?

Answer: We would *not* enter into a wildlife management contract with the farm. You would pay nothing even if you had offered to buy a contract-share.

A.2 Market Experiment Mechanism Descriptions

The elicitation mechanisms administered in the market experiments are described below.

The Provision Point with Money Back Guarantee and Proportional Rebate

The following text was used in the market experiment to describe the mechanism:

If the total of your group's offers is more than enough to cover the costs of the contract, we will pay the costs to implement the contract and refund any extra money offered. All extra funds received will be refunded to everyone in proportion to their share of the total offers we received. Making your highest possible offer increases your group's chance to succeed in implementing this contract. Remember that you will pay no more than the amount you offer, and it is possible that you would pay less.

The Uniform Price Auction The following text was used in the market experiment to describe the Uniform Price Auction:

If the total of your group's offers is more than enough to cover the costs of the contract, then we will calculate a "group price" so that everyone who pays ends up paying the same price. We will try to find a group price that divides the contract cost evenly across the maximum number of people, while still collecting enough money. If the group price is higher than your offer, you pay nothing and receive a 100% refund. If the group price is lower than your offer, you pay only the group price and we will refund any excess money offered above that price. If too few people offer enough money, so that it is impossible to determine such a group price,

the contract will not be implemented and you will pay nothing. Making your highest possible offer increases your group's chance to succeed in implementing this contract. Remember that you will pay no more than the amount you offer, and it is possible that you would pay less.

The Uniform Price Cap The following text was used in the market experiment to describe the Uniform Price Cap:

We are asking for your money now, but we will use it only if necessary. That is: We are asking everyone in your community group to contribute to a dedicated fund to buy a farm wildlife contract for the 2008 Bobolink nesting season. On April 30, if the fund contains sufficient money, we will buy the farm-wildlife contract. We will return any leftover money as follows. We will look for the lowest contribution that we can set as a "contribution cap" and still buy the contract. If your contribution was above this cap, we will return to you the amount you contributed over the cap. If the fund does not contain enough to pay for the contract, then we will return all money collected and the hayfield will not be managed for Bobolinks this year. This approach is designed to bring many people to participate at the same time, which means costs to you and each Jamestown resident in your group will be kept low.

The Pivotal Mechanism The following text was used in the market experiment to describe the Pivotal Mechanism:

If the total of your group's offers is more than enough to cover the costs of the contract, we will implement the contract and determine your payment as follows: If the total of everyone else's offers - not including yours - is higher than the amount needed to implement your group's contract, then we really don't need your money. Because everyone else's offers are enough, we will implement the contract and you will pay nothing. If the total of everyone else's offers is not enough to implement the contract, then your decision could be critical. If your offer raises the total offers high enough so we can implement the contract, then we need your money and we will collect the portion of your offered amount to meet the contract cost. (If the total offers including your still falls short, then we cannot implement the contract and we will refund your money.) Because you pay only when your decision is critical, it is in your interest to offer the highest amount you feel the farm-wildlife contract is worth to you. If you value the contract more than your offer, and if your decision is critical, a lower offer may prevent us from implementing the contract, when your highest value would have implemented the contract.

APPENDIX B

Convergence Plots and Descriptive Statistics for the Bayesian Analysis: Manuscript 2

B.1 The Models

Below are listed the final models for the second manuscript. V_{njt} denotes the value function of individual n for alternative j in choice situation t . A model that allows for heterogeneity in all parameters at the individual level would be specified as follows.

$$\begin{aligned}
 V_{njt} = & \alpha_n ASCBOTH_{njt} + \gamma_n ASCNO_{njt} + \sigma_n \beta_{A,n} Acres_{njt} + \sigma_n \beta_{R,n} Restore_{njt} + \\
 & \sigma_n \beta_{V,n} View_{njt} + \sigma_n \beta_{T,n} Tour_{njt} - \sigma_n cost_{njt} + \varepsilon_{njt}
 \end{aligned}
 \tag{B.1}$$

where the choice probabilities are

$$P_{ic} = \frac{e^{V_{ict}}}{\sum_j e^{V_{ijt}}}
 \tag{B.2}$$

The three models tested are listed below. R indicates that the parameter was estimated as a random parameter ($\beta_{x,n} = \beta_x + \eta_n$), while F indicates that the parameter was modeled as a fixed parameter.

Table B.1. Summary of Model Specifications

	Model 1	Model 2	Model 3
Price	F	R	R
No ASC	R	R	R
Both ASC	R	R	R
Acres	R	F	R
Restore	R	F	R
View	R	F	F
Tour	R	F	F

F: fixed coefficient, R: random coefficient

B.2 Convergence Indicators

There are several diagnostic tasks that can be employed to test for convergence. At the most basic level, the acceptance rate of the sampler can be monitored. We monitored and adjusted the acceptance rate to remain between 30-40% as suggested by Gelman et al. [142], who provide the rule of thumb when monitoring acceptance rates. Specifically, when $K = 1$, the optimal acceptance rate is about 0.44. This value drops to 0.23 as K rises. Lag-autocorrelations for the draws are supplied. High autocorrelations suggest slow mixing of the chains and slow convergence. Autocorrelations greater than 0.975 indicate that means and standard deviations of the chains are not reliable. In addition, trace plots, time series plots of the individual chains of the MCMC draws, were examined for stability. Kernel density plots for the population-level chains were examined as well. Well-defined shapes with clear maxima imply convergence. Finally, log plots which include time series plots of log-likelihood, root likelihood, parameter RMS, and average variance are supplied.

Autocorrelations and Lags

Table B.2. Model 1 Lags and Autocorrelations

	Lag 1	Lag 5	Lag 10	Lag 50
Price	0.202	0.101	0.057	0.020
Both ASC	0.278	0.155	0.121	0.060
No ASC	0.623	0.294	0.199	0.074
Acres ASC	0.738	0.406	0.254	0.029
Restore ASC	0.855	0.598	0.421	0.062
Tour ASC	0.996	0.991	0.988	0.963
View ASC	0.997	0.994	0.991	0.970

Table B.3. Model 2 Lags and Autocorrelations

	Lag 1	Lag 5	Lag 10	Lag 50
Price	0.393	0.107	0.034	0.013
Both ASC	0.348	0.114	0.078	0.047
No ASC	0.627	0.240	0.097	0.030
Acres WTP	0.297	0.070	0.027	0.017
Restore WTP	0.216	0.057	0.021	0.012
Tour WTP	0.985	0.930	0.868	0.515
View WTP	0.982	0.912	0.832	0.373

Table B.4. Model 3 Lags and Autocorrelations

	Lag 1	Lag 5	Lag 10	Lag 50
Price	0.403	0.135	0.078	0.006
Both ASC	0.614	0.446	0.351	0.097
No ASC	0.816	0.584	0.444	0.159
Acres WTP	0.842	0.608	0.443	0.073
Restore WTP	0.885	0.745	0.629	0.214
Tour WTP	0.201	0.061	0.046	0.006
View WTP	0.143	0.019	0.013	-0.005

Trace Plots

Figure B.1. Model 1 Trace Plot

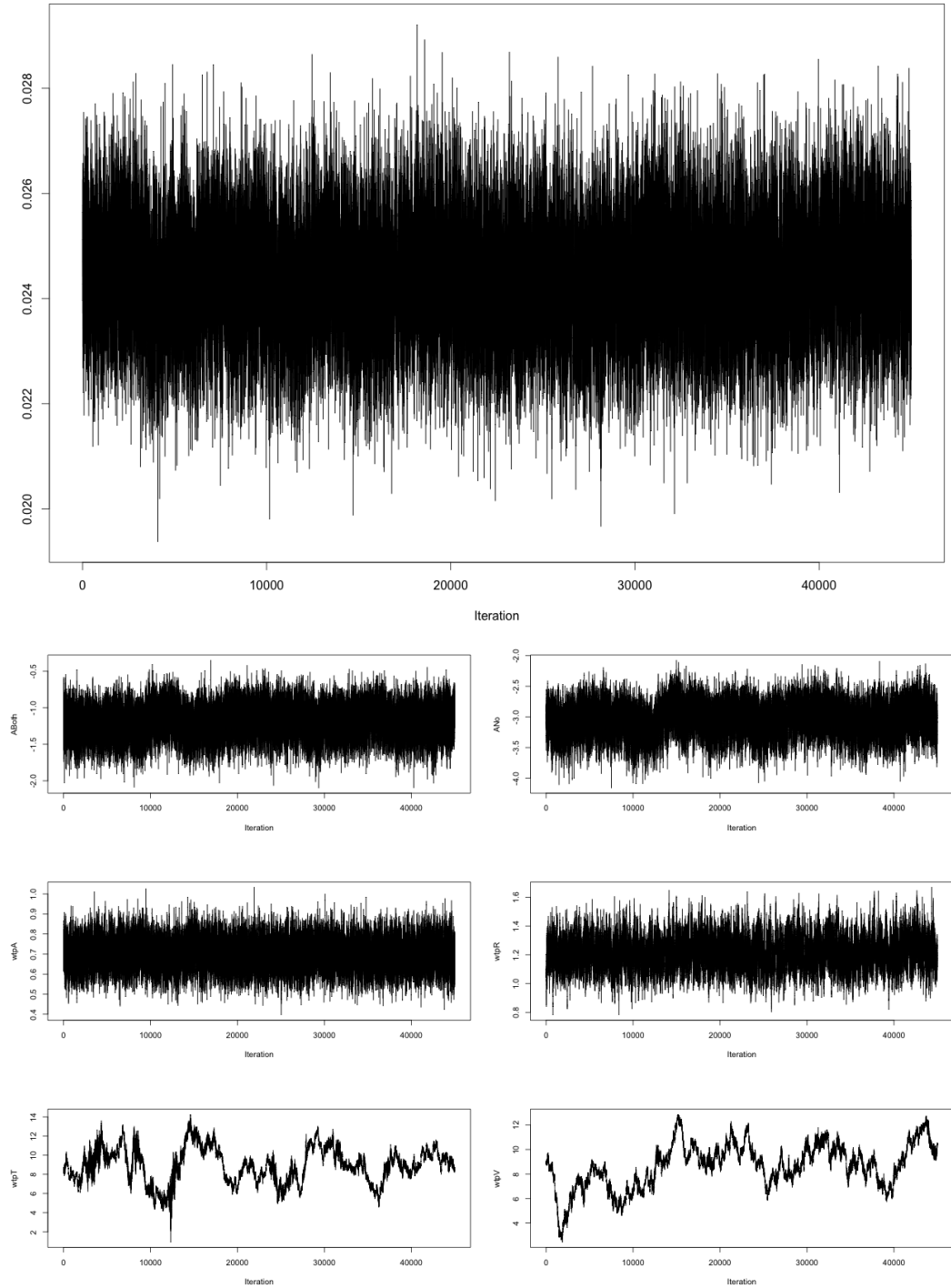


Figure B.2. Model 2 Trace Plot

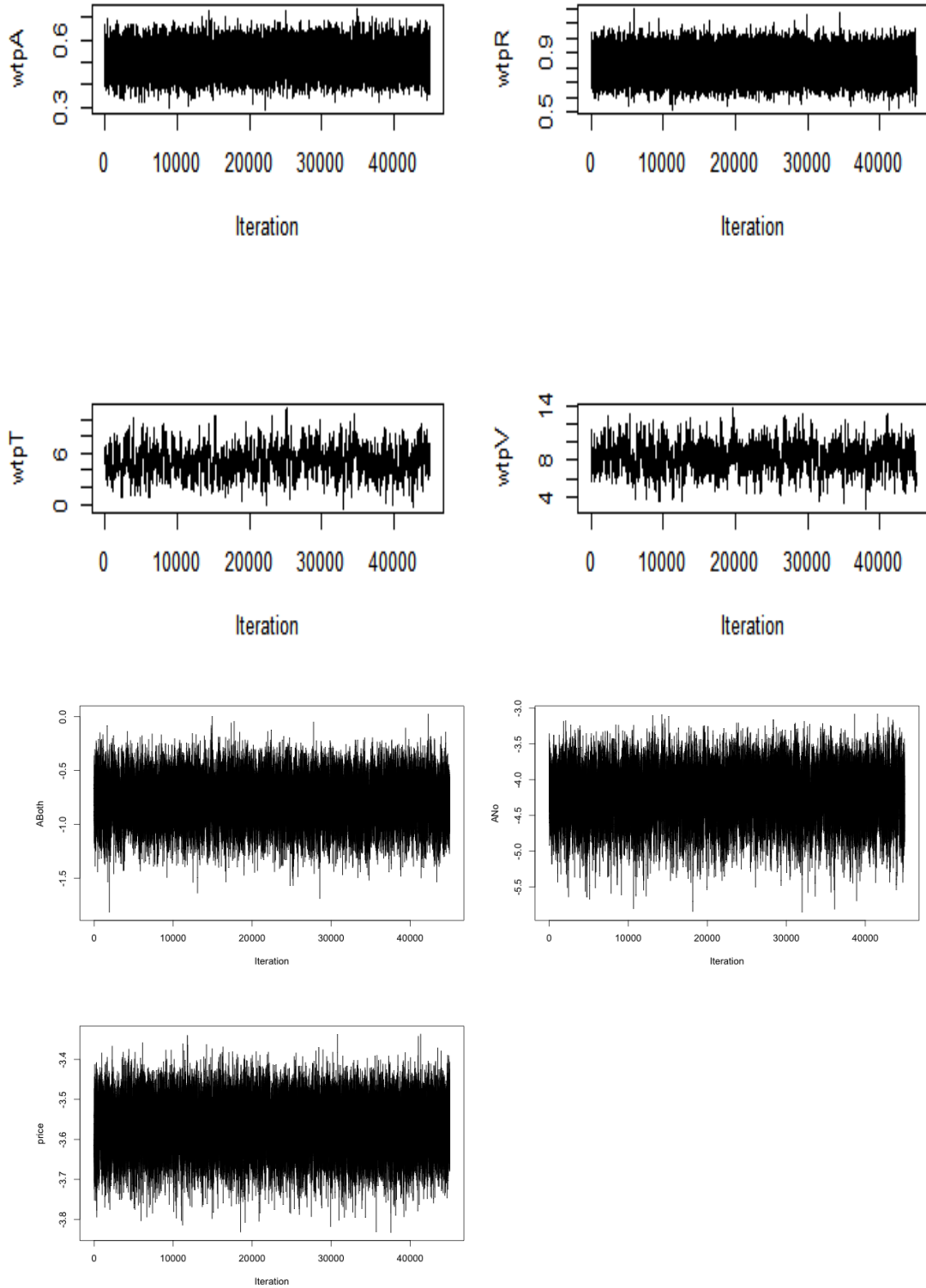
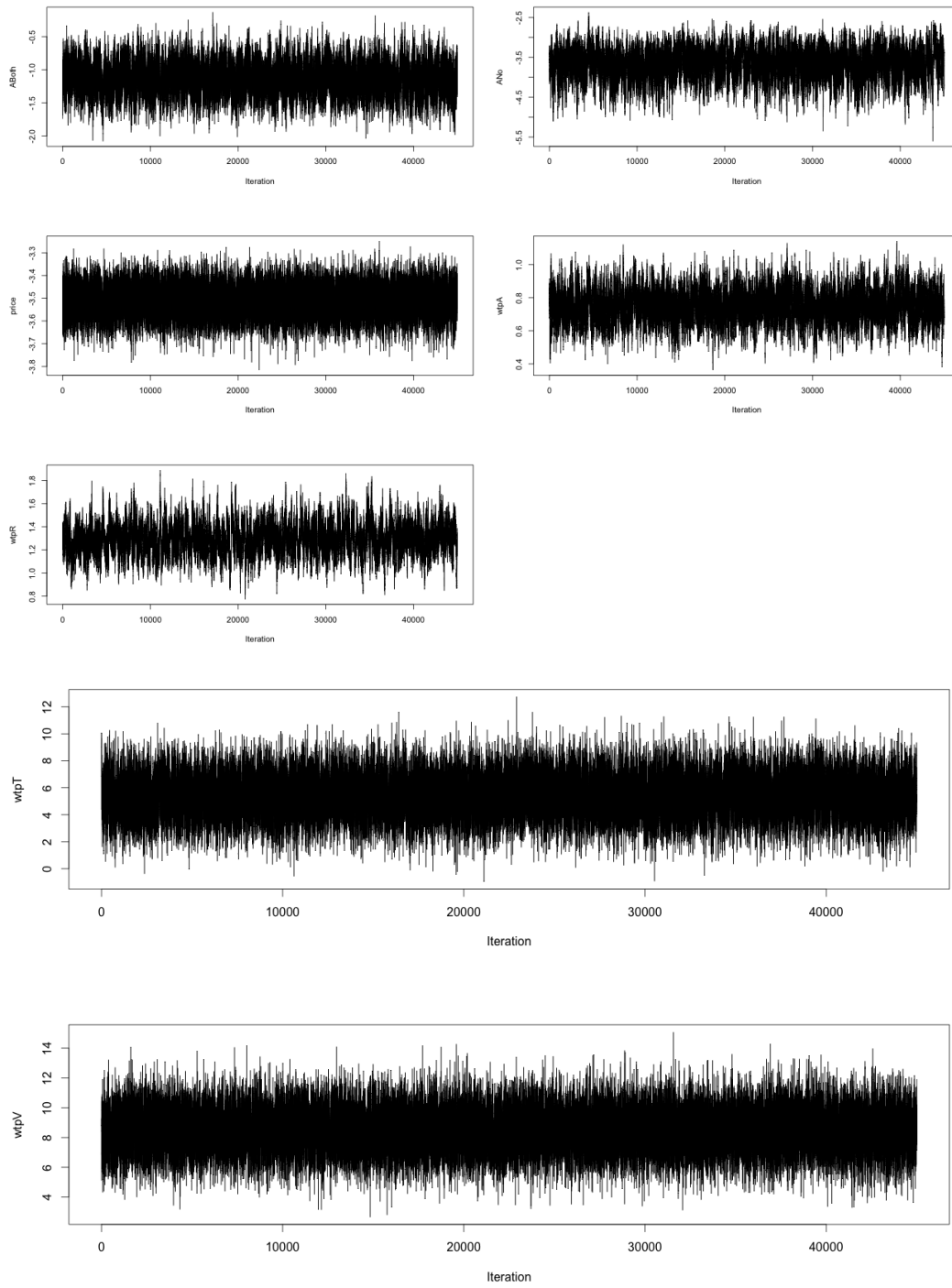


Figure B.3. Model 3 Trace Plot



Density Plots

Figure B.4. Model 1 Density Plot

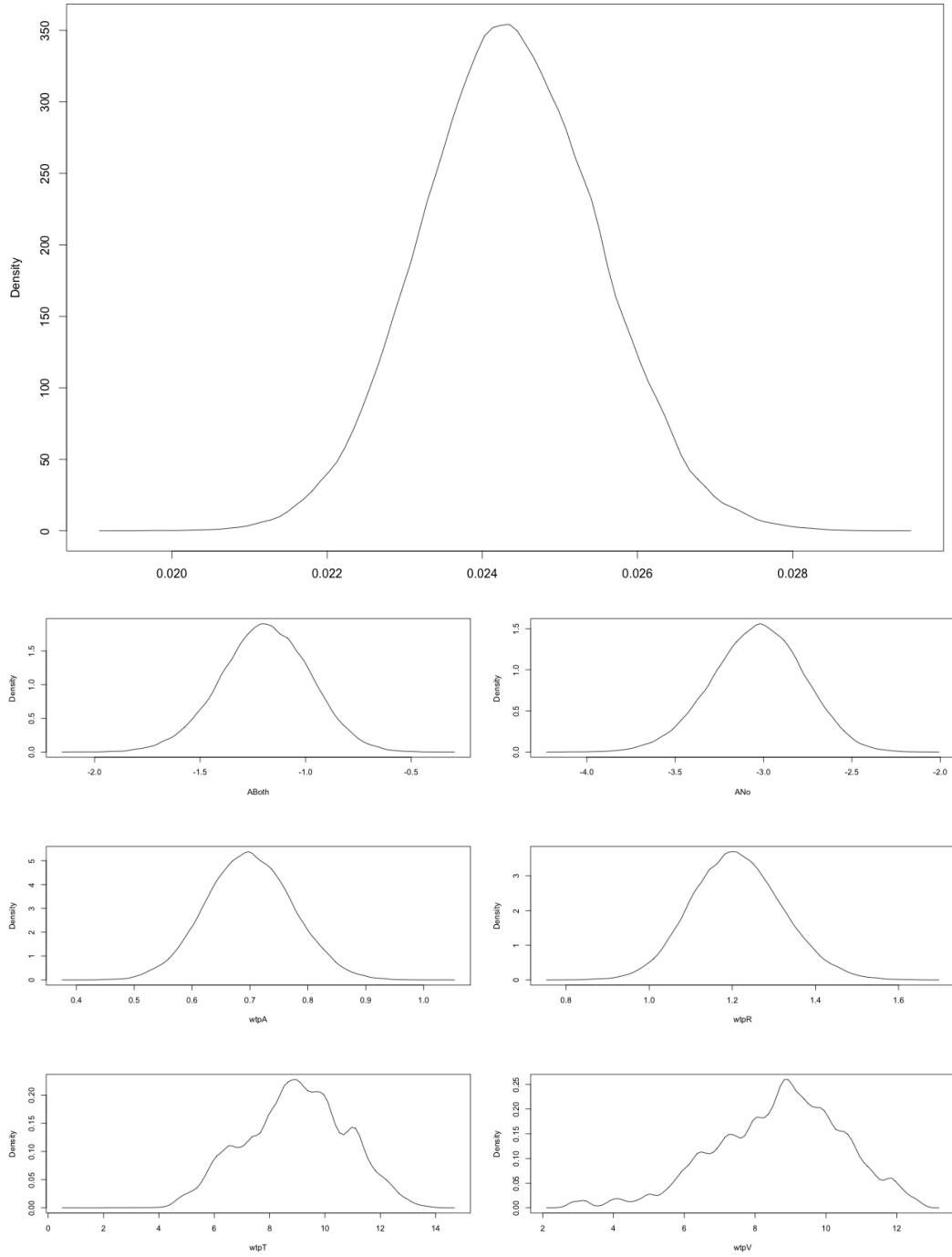


Figure B.5. Model 2 Density Plot

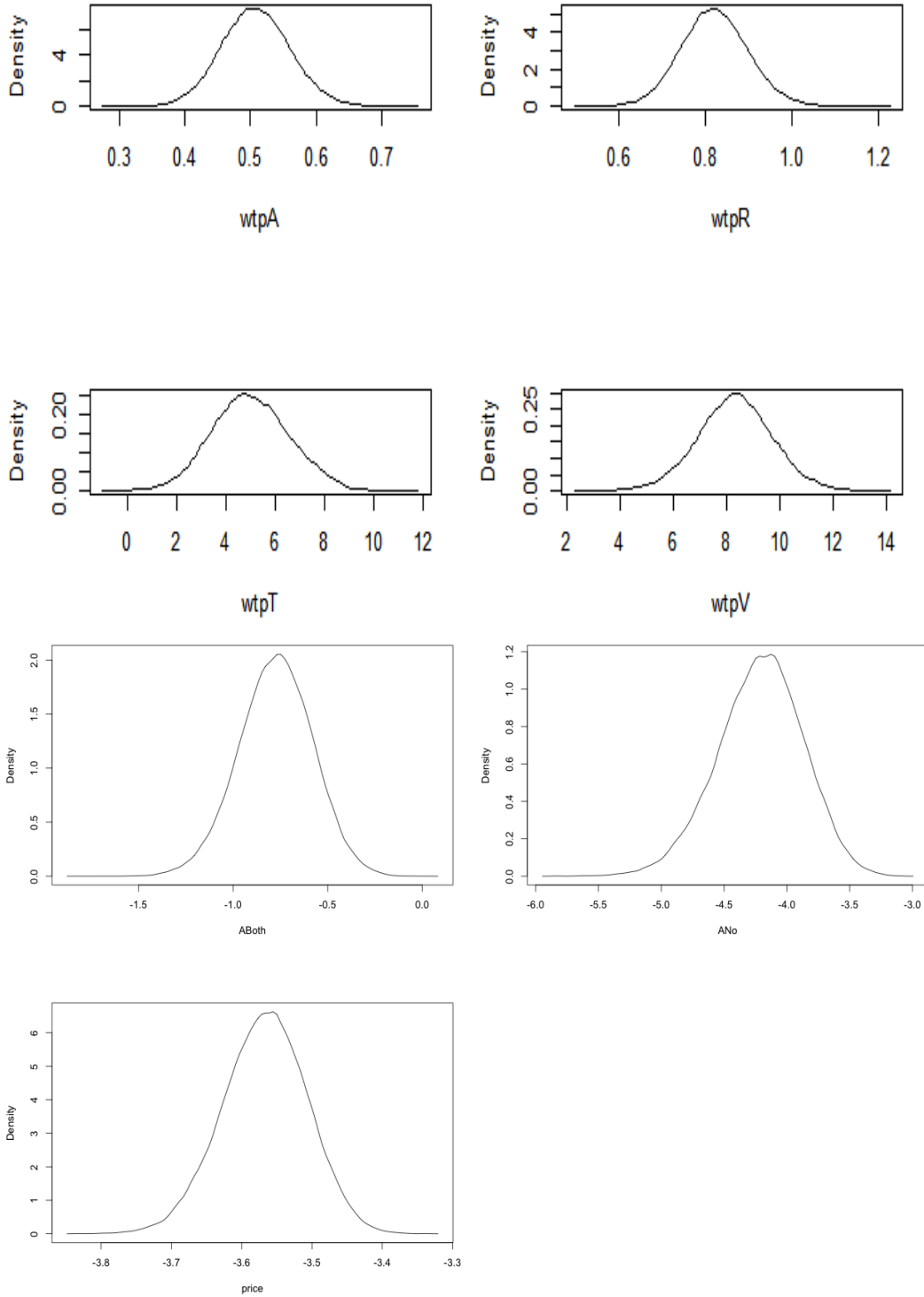
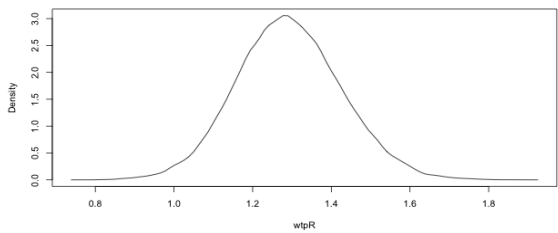
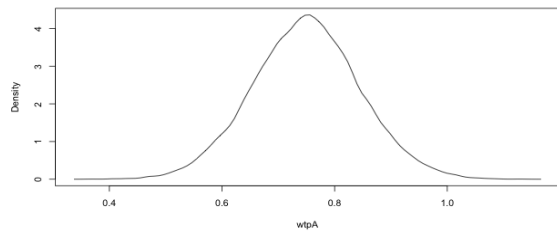
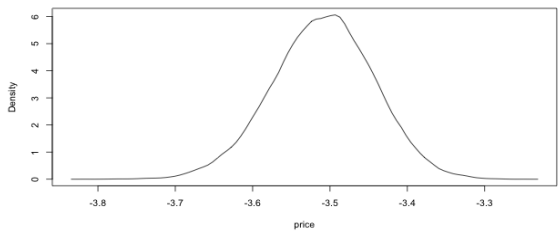
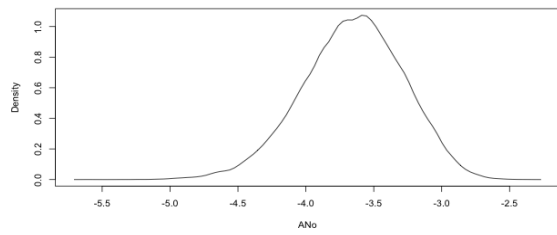
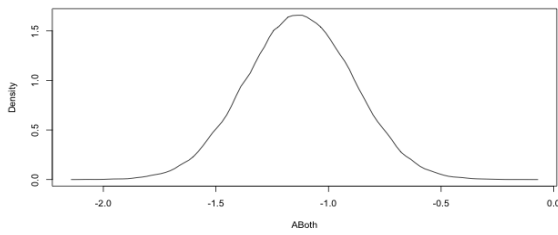
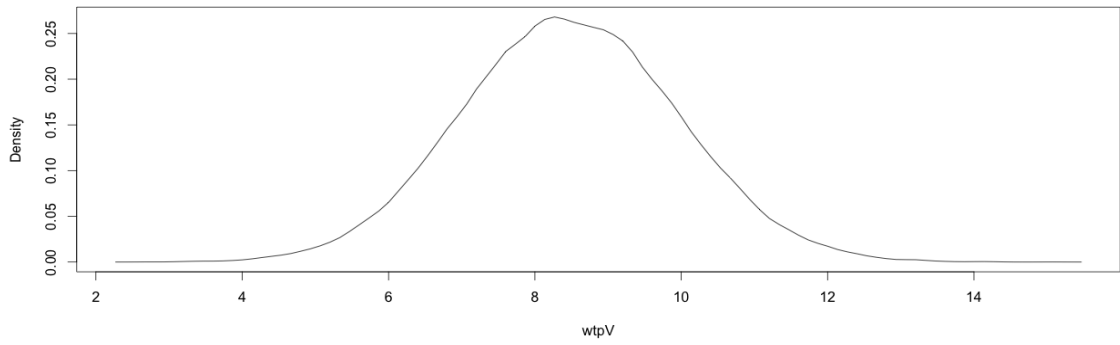
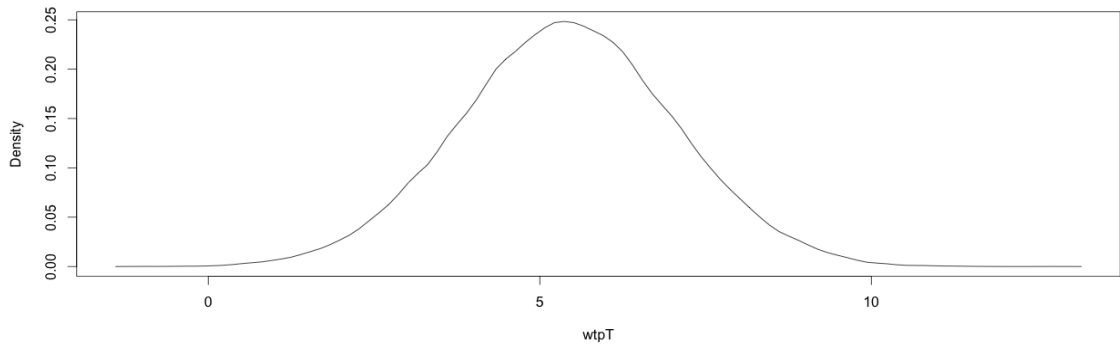


Figure B.6. Model 3 Density Plot



Log Plots

Figure B.7. Model 1 Log Plot

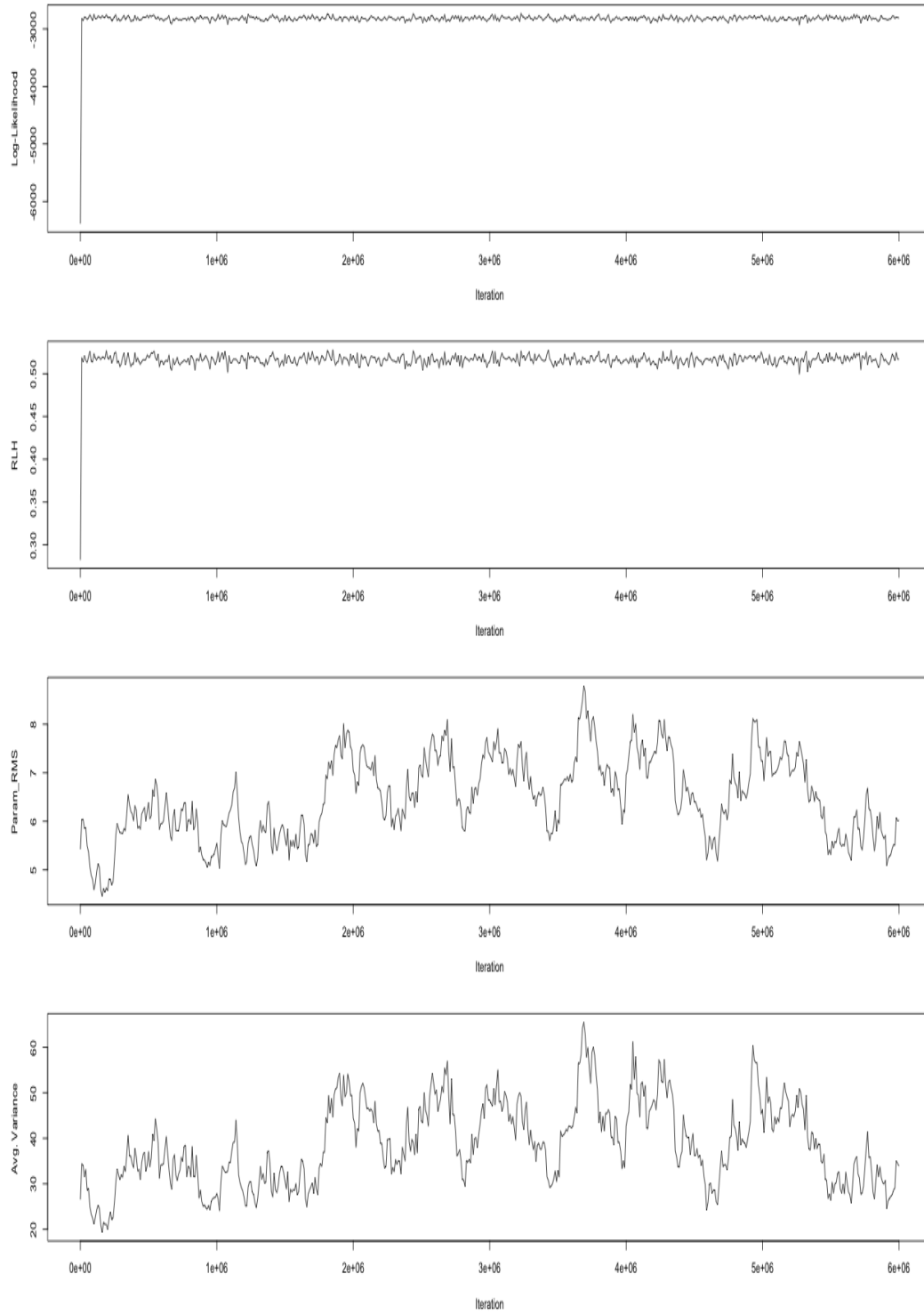


Figure B.8. Model 2 Log Plot

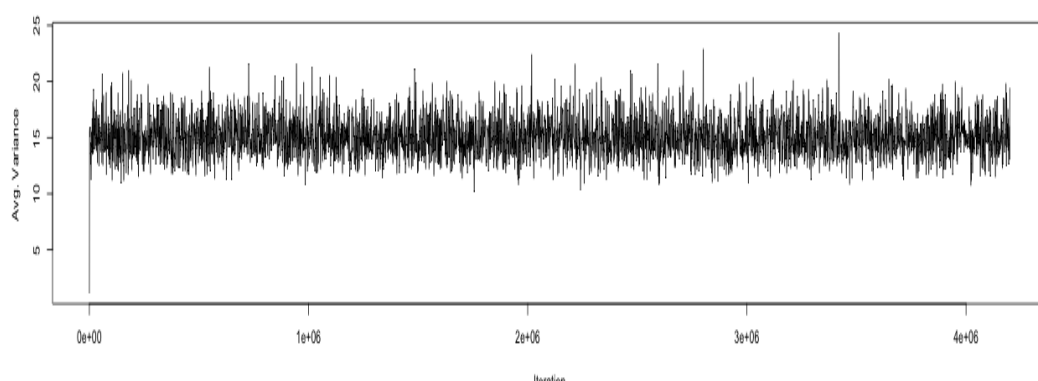
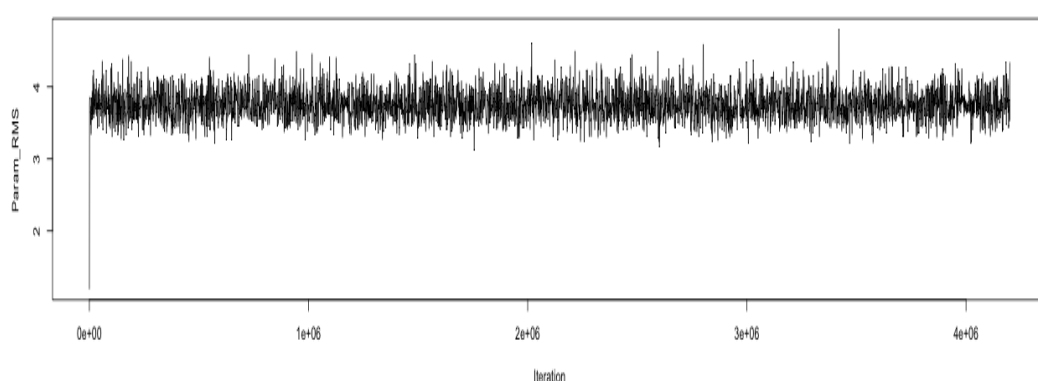
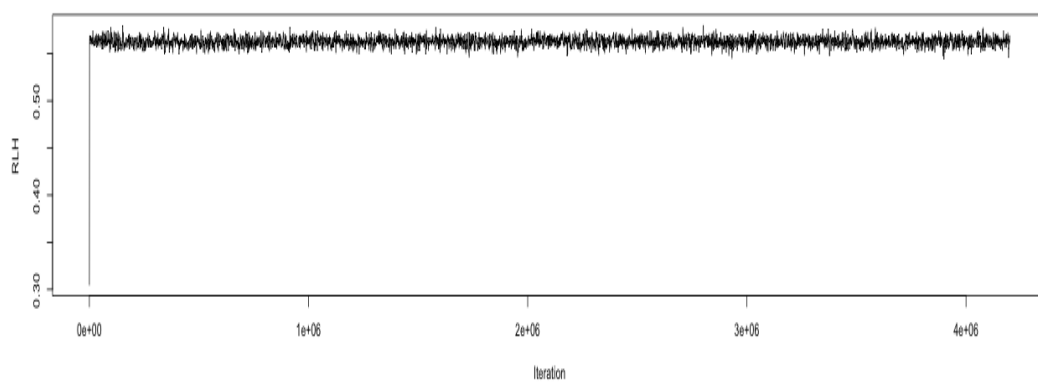
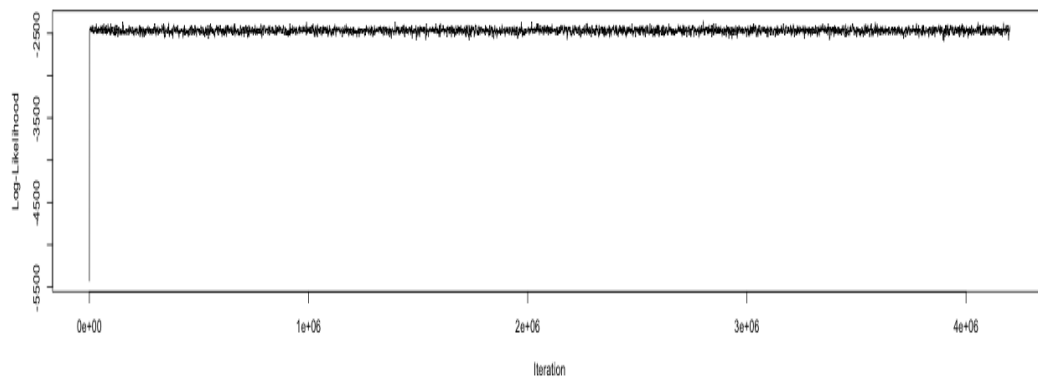
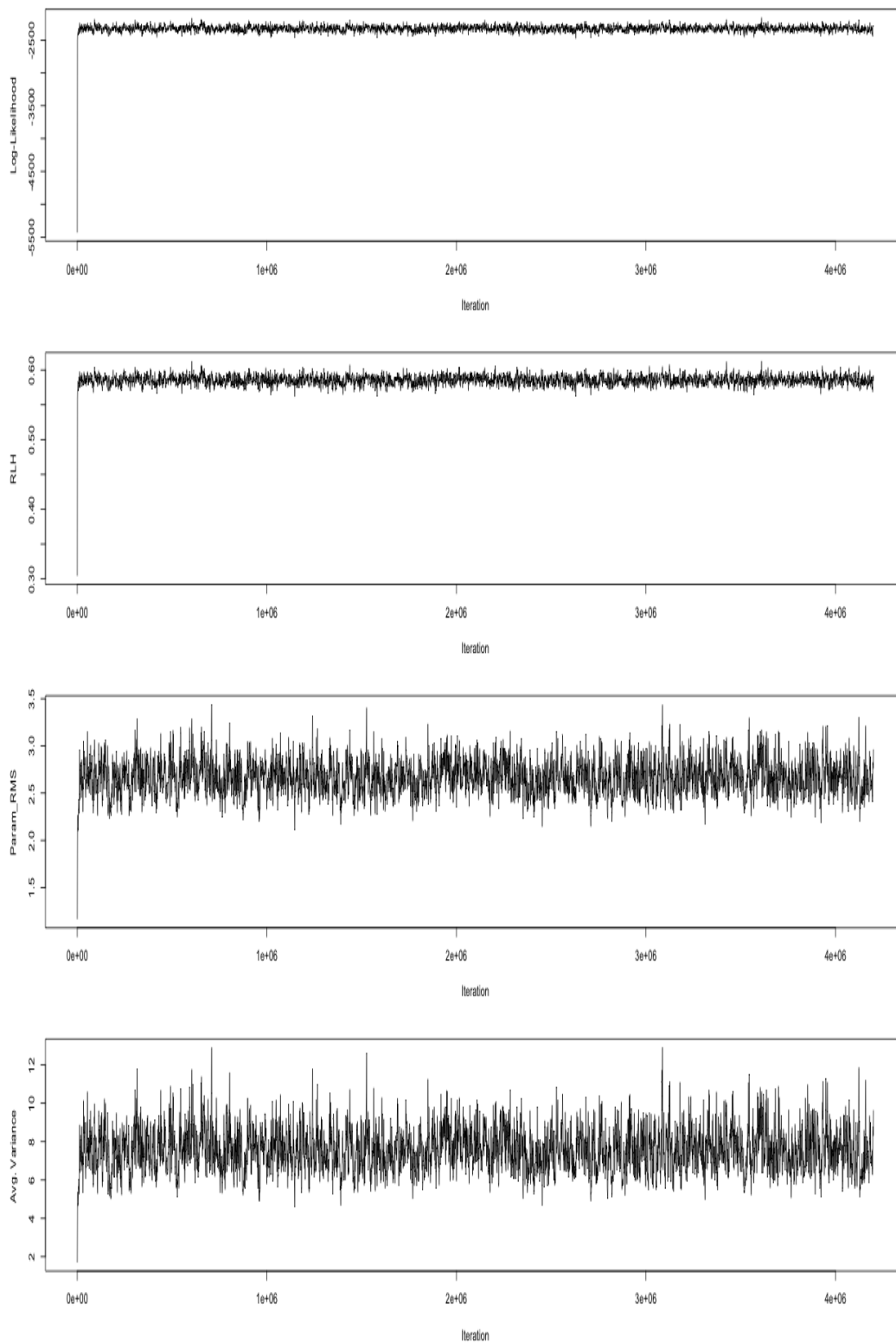


Figure B.9. Model 3 Log Plot



LIST OF REFERENCES

- [1] J. Hausman, “Contingent valuation: from dubious to hopeless,” *The Journal of Economic Perspectives*, vol. 26, no. 4, pp. 43–56, 2012.
- [2] T. C. Haab, M. G. Interis, D. R. Petrolia, and J. C. Whitehead, “From hopeless to curious? thoughts on hausman’s dubious to hopeless critique of contingent valuation,” *Applied Economic Perspectives and Policy*, vol. 35, no. 4, pp. 593–612, 2013.
- [3] S. Jaffry, H. Pickering, Y. Ghulam, D. Whitmarsh, and P. Wattage, “Consumer choices for quality and sustainability labelled seafood products in the uk,” *Food Policy*, vol. 29, no. 3, pp. 215–228, 2004.
- [4] C. R. Wessells, R. J. Johnston, and H. Donath, “Assessing consumer preferences for ecolabeled seafood: the influence of species, certifier, and household attributes,” *American Journal of Agricultural Economics*, vol. 81, no. 5, pp. 1084–1089, 1999.
- [5] R. J. Johnston, C. R. Wessells, H. Donath, and F. Asche, “Measuring consumer preferences for ecolabeled seafood: an international comparison,” *Journal of Agricultural and Resource Economics*, pp. 20–39, 2001.
- [6] R. J. Johnston and C. A. Roheim, “A battle of taste and environmental convictions for ecolabeled seafood: A contingent ranking experiment,” *Journal of Agricultural and Resource Economics*, pp. 283–300, 2006.
- [7] D. Brécard, B. Hlaimi, S. Lucas, Y. Perraudau, and F. Salladarré, “Determinants of demand for green products: An application to eco-label demand for fish in europe,” *Ecological economics*, vol. 69, no. 1, pp. 115–125, 2009.
- [8] G. L. Poe, J. E. Clark, D. Rondeau, and W. D. Schulze, “Provision point mechanisms and field validity tests of contingent valuation,” *Environmental and Resource Economics*, vol. 23, no. 1, pp. 105–131, 2002.
- [9] D. Rondeau, W. D. Schulze, and G. L. Poe, “Voluntary revelation of the demand for public goods using a provision point mechanism,” *Journal of Public Economics*, vol. 72, no. 3, pp. 455–470, 1999.
- [10] S. K. Rose, J. Clark, G. L. Poe, D. Rondeau, and W. D. Schulze, “The private provision of public goods: tests of a provision point mechanism for funding green power programs,” *Resource and energy economics*, vol. 24, no. 1, pp. 131–155, 2002.
- [11] P. Milgrom, *Putting auction theory to work*. Cambridge Univ Pr, 2004.

- [12] C. Das, *Investigating ways to improve discrete choice methods in assessing individual preference functions with greater reliability and accuracy*. ProQuest, 2007.
- [13] N. Hanley, R. E. Wright, and V. Adamowicz, “Using choice experiments to value the environment,” *Environmental and resource economics*, vol. 11, no. 3-4, pp. 413–428, 1998.
- [14] N. Hanley, R. E. Wright, and B. Alvarez-Farizo, “Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive,” *Journal of environmental management*, vol. 78, no. 2, pp. 183–193, 2006.
- [15] N. Hanley, R. E. Wright, and G. Koop, “Modelling recreation demand using choice experiments: climbing in Scotland,” *Environmental and Resource Economics*, vol. 22, no. 3, pp. 449–466, 2002.
- [16] C.-Z. Li, J. Kuuluvainen, E. Pouta, M. Rekola, and O. Tahvonen, “Using choice experiments to value the Natura 2000 nature conservation programs in Finland,” *Environmental and Resource Economics*, vol. 29, no. 3, pp. 361–374, 2004.
- [17] W. Adamowicz, P. Boxall, M. Williams, and J. Louviere, “Stated preference approaches for measuring passive use values: choice experiments and contingent valuation,” *American journal of agricultural economics*, vol. 80, no. 1, pp. 64–75, 1998.
- [18] F. Carlsson and P. Martinsson, “Do hypothetical and actual marginal willingness to pay differ in choice experiments?: Application to the valuation of the environment,” *Journal of Environmental Economics and Management*, vol. 41, no. 2, pp. 179–192, 2001.
- [19] S. Colombo, N. Hanley, and J. Calatrava-Requena, “Designing policy for reducing the off-farm effects of soil erosion using choice experiments,” *Journal of Agricultural Economics*, vol. 56, no. 1, pp. 81–95, 2005.
- [20] S. Caussade, J. d. D. Ortúzar, L. I. Rizzi, and D. A. Hensher, “Assessing the influence of design dimensions on stated choice experiment estimates,” *Transportation research part B: Methodological*, vol. 39, no. 7, pp. 621–640, 2005.
- [21] K. Sælensminde, “Inconsistent choices in stated choice data; use of the logit scaling approach to handle resulting variance increases,” *Transportation*, vol. 28, no. 3, pp. 269–296, 2001.
- [22] J. DeShazo and G. Fermo, “Designing choice sets for stated preference methods: the effects of complexity on choice consistency,” *Journal of Environmental Economics and management*, vol. 44, no. 1, pp. 123–143, 2002.

- [23] W. Adamowicz, D. Bunch, T. Cameron, B. Dellaert, M. Hanneman, M. Keane, J. Louviere, R. Meyer, T. Steenburgh, and J. Swait, “Behavioral frontiers in choice modeling,” *Marketing Letters*, vol. 19, pp. 215–228, 2008, 10.1007/s11002-008-9038-1. [Online]. Available: <http://dx.doi.org/10.1007/s11002-008-9038-1>
- [24] K. Train and M. Weeks, “Discrete choice models in preference space and willingness-to-pay space,” in *Applications of Simulation Methods in Environmental and Resource Economics*, A. Alberini and R. Scarpa, Eds. Boston, Massachusetts: Kluwer Academic Publishers, 2005, ch. 1, pp. 1–16.
- [25] T. A. Cameron and J. DeShazo, “Differential attention to attributes in utility-theoretic choice models,” *Journal of choice modelling*, vol. 3, no. 3, pp. 73–115, 2010.
- [26] D. A. Hensher and J. M. Rose, “Simplifying choice through attribute preservation or non-attendance: implications for willingness to pay,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 4, pp. 583–590, 2009.
- [27] S. Hess and D. A. Hensher, “Using conditioning on observed choices to retrieve individual-specific attribute processing strategies,” *Transportation Research Part B: Methodological*, vol. 44, no. 6, pp. 781–790, 2010.
- [28] D. A. Hensher, J. M. Rose, and W. H. Greene, “Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design,” *Transportation*, vol. 39, no. 2, pp. 235–245, 2012.
- [29] M. D. Smith, C. A. Roheim, L. B. Crowder, B. S. Halpern, M. Turnipseed, J. L. Anderson, F. Asche, L. Bourillón, A. G. Guttormsen, A. Khan, *et al.*, “Sustainability and global seafood,” *Science*, vol. 327, no. 5967, pp. 784–786, 2010.
- [30] O. o. P. P. U.S. Environmental Protection Agency (US EPA) and Toxics, “Environmental labeling: Issues, policies, and practices worldwide,” Office of Prevention, Pesticides and Toxic Substances, Washington, DC, Tech. Rep. 742-R-98-009, December 1998.
- [31] E. Gudmundsson and C. R. Wessells, “Ecolabeling seafood for sustainable production: implications for fisheries management,” *Marine Resource Economics*, vol. 15, no. 2, pp. 97–113, 2000.
- [32] R. A. Sedjo and S. K. Swallow, “Voluntary eco-labeling and the price premium,” *Land economics*, vol. 78, no. 2, pp. 272–284, 2002.

- [33] M. F. Teisl, S. Peavey, F. Newman, J. Buono, and M. Hermann, “Consumer reactions to environmental labels for forest products: A preliminary look,” *Forest Products Journal*, vol. 52, no. 1, pp. 44–50, 2002.
- [34] B. Cashore, G. Auld, and D. Newsom, “Forest certification (eco-labeling) programs and their policy-making authority: explaining divergence among North American and European case studies,” *Forest Policy and Economics*, vol. 5, no. 3, pp. 225–247, 2003.
- [35] C. A. Roheim *et al.*, “The economics of ecolabelling,” *Seafood Ecolabelling: Principles and Practice*, pp. 38–57, 2008.
- [36] G. Parkes, J. A. Young, S. F. Walmsley, R. Abel, J. Harman, P. Horvat, A. Lem, A. MacFarlane, M. Mens, and C. Nolan, “Behind the signs—a global review of fish sustainability information schemes,” *Reviews in Fisheries Science*, vol. 18, no. 4, pp. 344–356, 2010.
- [37] F. Salladarré, P. Guillotreau, Y. Perraudeau, and M.-C. Monfort, “The demand for seafood eco-labels in France,” *Journal of Agricultural & Food Industrial Organization*, vol. 8, no. 1, 2010.
- [38] L. K. Ozanne and R. P. Vlosky, “Willingness to pay for environmentally certified wood products: A consumer perspective,” *Forest Products Journal*, vol. 47, no. 6, pp. 39–48, 1997.
- [39] L. K. Ozanne and R. P. Vlosky, “Certification from the US consumer perspective: a comparison of 1995 and 2000,” *Forest Products Journal*, vol. 53, no. 3, pp. 13–21, 2003.
- [40] K. Forsyth, D. Haley, and R. Kozak, “Will consumers pay more for certified wood products?” *Journal of Forestry*, vol. 97, no. 2, pp. 18–22, 1999.
- [41] F. X. Aguilar and R. P. Vlosky, “Consumer willingness to pay price premiums for environmentally certified wood products in the US,” *Forest Policy and Economics*, vol. 9, no. 8, pp. 1100–1112, 2007.
- [42] O. H. S. T. Force. “Closing the net: Stopping illegal fishing on the high seas.” 2009. [Online]. Available: <http://www.high-seas.org>
- [43] S. Washington, “Ecolabels and marine capture fisheries: Current practice and emerging issues,” *GLOBEFISH Research Programme*, vol. 91, 2008.
- [44] S. Rosen, “Hedonic prices and implicit markets: product differentiation in pure competition,” *The Journal of Political Economy*, vol. 82, no. 1, pp. 34–55, 1974.
- [45] I. Galarraga and A. Markandya, “Economic techniques to estimate the demand for sustainable products: A case study for fair trade and organic coffee in the United Kingdom,” *Economía Agraria y Recursos Naturales*, vol. 4, no. 7, pp. 109–134, 2004.

- [46] B.-H. Lin, T. A. Smith, and C. L. Huang, "Organic premiums of US fresh produce," *Renewable Agriculture and Food Systems*, vol. 23, no. 3, pp. 208–216, 2008.
- [47] T. Smith, C. Huang, and B.-H. Lin, "Estimating organic premiums in the US fluid milk market," *Renewable agriculture and food systems*, vol. 24, no. 03, pp. 197–204, 2009.
- [48] F. Zhang, J. E. Epperson, C. L. Huang, and J. E. Houston, "Organic price premiums paid for fresh tomatoes and apples by US households: Evidence from Nielsen homescan data," *Journal of Food Distribution Research*, vol. 40, no. 3, pp. 105–114, 2009.
- [49] M. S. Council, "Marine Stewardship Council annual report 2009/ 2010, 2009." [Online]. Available: http://www.msc.org/documents/msc-brochures/annual-report-archive/MSC_AnnualReview_final_web.pdf/view
- [50] N. M. F. S. (NMFS), "Fisheries of the United States, 2008," Fisheries Statistics Division, Silver Spring, Maryland, Tech. Rep., 2009.
- [51] L. H. Gulbrandsen, "Creating markets for eco-labelling: are consumers insignificant?" *International Journal of Consumer Studies*, vol. 30, no. 5, pp. 477–489, 2006.
- [52] K. A. O'Brien and M. F. Teisl, "Eco-information and its effect on consumer values for environmentally certified forest products," *Journal of Forest Economics*, vol. 10, no. 2, pp. 75–96, 2004.
- [53] C. A. Roheim, "Early indications of market impacts from the Marine Stewardship Council's ecolabeling of seafood," *Marine resource economics*, vol. 18, no. 1, 2003.
- [54] C. Roheim and J. Sutinen, "Trade and market-related instruments to reinforce fisheries management measures to promote sustainable fishing practices, report prepared for the International Centre for Trade and Sustainable Development (ICTSD) and the High Seas Task Force," Available at www.high-seas.org, 2006.
- [55] R. W. Cotterill, "Scanner data: New opportunities for demand and competitive strategy analysis," *Agricultural and Resource Economics Review*, vol. 23, no. 2, 1994.
- [56] M. Christen, S. Gupta, J. C. Porter, R. Staelin, and D. R. Wittink, "Using market-level data to understand promotion effects in a nonlinear model," *Journal of Marketing Research*, pp. 322–334, 1997.
- [57] H. Theil, *Principles of Econometrics*. Wiley, 1971.

- [58] J. L. Lusk, “The effect of Proposition 2 on the demand for eggs in California,” *Journal of Agricultural & Food Industrial Organization*, vol. 8, no. 1, 2010.
- [59] R. W. Cotterill, W. P. Putsis, Jr, and R. Dhar, “Assessing the competitive interaction between private labels and national brands,” *The Journal of Business*, vol. 73, no. 1, pp. 109–137, 2000.
- [60] R. W. Cotterill and W. P. Putsis Jr, “Market share and price setting behavior for private labels and national brands,” *Review of Industrial Organization*, vol. 17, no. 1, pp. 17–39, 2000.
- [61] K. E. McConnell and I. E. Strand, “Hedonic prices for fish: tuna prices in Hawaii,” *American Journal of Agricultural Economics*, vol. 82, no. 1, pp. 133–144, 2000.
- [62] M. T. Carroll, J. L. Anderson, and J. Martínez-Garmendia, “Pricing US North Atlantic bluefin tuna and implications for management,” *Agribusiness*, vol. 17, no. 2, pp. 243–254, 2001.
- [63] G. Sogn-Grundvåg, T. A. Larsen, and J. A. Young, “The value of line-caught and other attributes: an exploration of price premiums for chilled fish in UK supermarkets,” *Marine Policy*, vol. 38, pp. 41–44, 2013.
- [64] F. Asche, T. A. Larsen, M. D. Smith, G. Sogn-Grundvåg, and J. A. Young, “Pricing of eco-labels for salmon in UK supermarkets,” Duke Working Paper EE 13-02, downloaded from: <http://sites.nicholasinstitute.duke.edu/environmentaleconomics/files/2013/05/WP-EE-13-02.pdf> (17/09/2013), Tech. Rep., 2012.
- [65] C. A. Roheim, L. Gardiner, and F. Asche, “Value of brands and other attributes: Hedonic analysis of retail frozen fish in the UK,” *Marine Resource Economics*, vol. 22, no. 3, p. 239, 2007.
- [66] W. Nimon and J. Beghin, “Are eco-labels valuable? evidence from the apparel industry,” *American Journal of Agricultural Economics*, vol. 81, no. 4, pp. 801–811, 1999.
- [67] H. White, “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica: Journal of the Econometric Society*, pp. 817–838, 1980.
- [68] J. G. MacKinnon and H. White, “Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties,” *Journal of Econometrics*, vol. 29, no. 3, pp. 305–325, 1985.
- [69] R. Davidson and J. G. MacKinnon, *Econometric theory and methods*. Oxford University Press New York, 2004.

- [70] J. M. Wooldridge, *Introductory econometrics: a modern approach*. Ohio: Thomson South-Western Press, 2003.
- [71] J. Andreoni, “Impure altruism and donations to public goods: a theory of warm-glow giving,” *The economic journal*, vol. 100, no. 401, pp. 464–477, 1990.
- [72] E. K. Yiridoe, S. Bonti-Ankomah, and R. C. Martin, “Comparison of consumer perceptions and preference toward organic versus conventionally produced foods: a review and update of the literature,” *Renewable Agriculture and Food Systems*, vol. 20, no. 4, pp. 193–205, 2005.
- [73] J. Gilmore, “Case study 3: MSC certification of the Alaska pollock fishery,” *Seafood Ecolabelling: Principles and Practice*, pp. 269–286, 2008.
- [74] A. Kinzig, C. Perrings, F. Chapin, S. Polasky, V. Smith, D. Tilman, and B. Turner, “Paying for ecosystem services—promise and peril,” *Science*, vol. 334, no. 6056, pp. 603–604, 2011.
- [75] S. K. Swallow, E. C. Smith, E. Uchida, and C. M. Anderson, “Ecosystem services beyond valuation, regulation and philanthropy: Integrating consumer values into the economy,” *Choices*, vol. 23, no. 2, 2008. [Online]. Available: <http://EconPapers.repec.org/RePEc:ags:aaeach:94656>
- [76] E. Uchida, C. M. Anderson, and S. K. Swallow, “Marketing ecosystem services from agricultural land: Stated preferences over payment mechanisms and actual sales of farm-wildlife contracts,” American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association), 2007 Annual Meeting, July 29-August 1, 2007, Portland, Oregon TN 9955, 2007. [Online]. Available: <http://ideas.repec.org/p/ags/aaea07/9955.html>
- [77] C. A. Crystal Fry, Emi Uchida and S. Swallow, “Summary report: Working farms and wildlife in Jamestown, RI: 2006 survey of Jamestown residents,” University of Rhode Island, Kingston, RI, Tech. Rep., 2007.
- [78] B. G. Dellaert, J. D. Brazell, and J. J. Louviere, “The effect of attribute variation on consumer choice consistency,” *Marketing Letters*, vol. 10, no. 2, pp. 139–147, 1999.
- [79] B. Halvorsen and K. Sølensminde, “Differences between willingness-to-pay estimates from open-ended and discrete-choice contingent valuation methods: the effects of heteroscedasticity,” *Land Economics*, pp. 262–282, 1998.
- [80] T. A. Cameron, G. L. Poe, R. G. Ethier, and W. D. Schulze, “Alternative non-market value-elicitation methods: are the underlying preferences the same?” *Journal of Environmental Economics and Management*, vol. 44, no. 3, pp. 391–425, 2002.

- [81] J. Louviere and T. Eagle, “Confound it! that pesky little scale constant messes up our convenient assumptions,” in *Proceedings 2006 Sawtooth Software Conference*. Sequem, WA: Sawtooth Software, 2006, pp. 211–228.
- [82] L. Mitchell, C. R. Smith, and R. A. Malecki, *Ecology of grassland breeding birds in the northeastern United States: a literature review with recommendations for management*. New York Cooperative Fish and Wildlife Research Unit, Department, of Natural Resources, Cornell University, 2000.
- [83] C. F. Camerer and E. Fehr, “When does “economic man” dominate social behavior?” *Science*, vol. 311, no. 5757, pp. 47–52, 2006.
- [84] K. J. Lancaster, “A new approach to consumer theory,” *Journal of Political Economy*, vol. 74, pp. 132–157, 1966.
- [85] S. Rosen, “Hedonic prices and implicit markets: Product differentiation in pure competition,” *Journal of Political Economy*, vol. 82, pp. 34–55, 1974.
- [86] F. Carlsson, P. Frykblom, and C. Liljenstolpe, “Valuing wetland attributes: an application of choice experiments,” *Ecological Economics*, vol. 47, no. 1, pp. 95 – 103, 2003. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0921800903002301>
- [87] C. R. Bhat, “Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model,” *Transportation Research Part B: Methodological*, vol. 35, no. 7, pp. 677 – 693, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S019126150000014X>
- [88] M. Sillano and J. d. D. Ortúzar, “Willingness-to-pay estimation with mixed logit models: some new evidence,” *Environment and Planning A*, vol. 37, no. 3, pp. 525–550, 2005.
- [89] K. Train and G. Sonnier, “Mixed logit with bounded distributions of part-worths,” *Applications of Simulation Methods in Environmental Resource Economics*, edited by A. Alberini and R. Scarpa. New York: Kluwer Academic, 2003.
- [90] R. A. Daziano and E. Chiew, “On the effect of the prior of Bayes estimators of the willingness to pay for electric-vehicle driving range,” *Transportation Research Part D: Transport and Environment*, vol. 21, pp. 7–13, 2013.
- [91] C. Das, C. M. Anderson, and S. K. Swallow, “Estimating distributions of willingness to pay for heterogeneous populations,” *Southern Economic Journal*, pp. 593–610, 2009.
- [92] K. Balcombe, A. Chalak, and I. Fraser, “Model selection for the mixed logit with Bayesian estimation,” *Journal of Environmental Economics and*

- Management*, vol. 57, no. 2, pp. 226 – 237, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0095069608000739>
- [93] R. Scarpa, M. Thiene, and K. Train, “Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps,” *American Journal of Agricultural Economics*, vol. 90, no. 4, pp. 994–1010, 2008. [Online]. Available: <http://ajae.oxfordjournals.org/content/90/4/994.abstract>
- [94] R. Scarpa, M. Thiene, and D. A. Hensher, “Preferences for tap water attributes within couples: An exploration of alternative mixed logit parameterizations,” *Water Resources Research*, vol. 48, no. 1, 2012.
- [95] K. Jedidi, S. Jagpal, and P. Manchanda, “Measuring heterogeneous reservation prices for product bundles,” *Marketing Science*, vol. 22, pp. 107–130, 2003.
- [96] D. G. Fiebig, M. P. Keane, J. Louviere, and N. Wasi, “The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity,” *Marketing Science*, vol. 29, no. 3, pp. 393–421, May 2010. [Online]. Available: <http://dx.doi.org/10.1287/mksc.1090.0508>
- [97] G. Sonnier, A. Ainslie, and T. Otter, “Heterogeneity distributions of willingness-to-pay in choice models,” *Quantitative Marketing and Economics*, vol. 5, pp. 313–331, 2007, 10.1007/s11129-007-9024-6. [Online]. Available: <http://dx.doi.org/10.1007/s11129-007-9024-6>
- [98] K. Train, *Discrete choice methods with simulation*. Cambridge university press, 2003.
- [99] P. E. Rossi and G. M. Allenby, “Bayesian statistics and marketing,” *Marketing Science*, vol. 22, pp. 304–328, 2003.
- [100] K. Train and G. Sonnier, “Mixed logit with bounded distributions of correlated partworths,” in *Applications of simulation methods in environmental and resource economics*. Springer, 2005, pp. 117–134.
- [101] P. E. Rossi, G. M. Allenby, and R. E. McCulloch, *Bayesian statistics and marketing*. J. Wiley & Sons, 2005.
- [102] B. Orme, “The CBC/HB system for hierarchical Bayes estimation version 5.0 technical paper,” *Technical Paper Series. Sawtooth Software*, 2009.
- [103] T. A. Cameron and J. Englin, “Welfare effects of changes in environmental quality under individual uncertainty about use,” *Rand Journal of Economics*, pp. S45–S70, 1997.
- [104] C. A. Vossler, J. Kerkvliet, S. Polasky, and O. Gainutdinova, “Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon,” *Journal of Economic Behavior & Organization*, vol. 51, no. 2, pp. 261–277, 2003.

- [105] L. Nahuelhual, M. L. Loureiro, and J. Loomis, “Using random parameters to account for heterogeneous preferences in contingent valuation of public open space,” *Journal of Agricultural and Resource Economics*, pp. 537–552, 2004.
- [106] B. Provencher, D. J. Lewis, and K. Anderson, “Disentangling preferences and expectations in stated preference analysis with respondent uncertainty: The case of invasive species prevention,” *Journal of Environmental Economics and Management*, 2012.
- [107] R. Scarpa, M. Thiene, and D. A. Hensher, “Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: does it matter?” *Land economics*, vol. 86, no. 4, pp. 817–839, 2010.
- [108] J. C. Whitehead, S. K. Pattanayak, G. L. Van Houtven, and B. R. Gelso, “Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science,” *Journal of Economic Surveys*, vol. 22, no. 5, pp. 872–908, 2008.
- [109] S. B. Caudill, P. A. Groothuis, and J. C. Whitehead, “The development and estimation of a latent choice multinomial logit model with application to contingent valuation,” *American Journal of Agricultural Economics*, vol. 93, no. 4, pp. 983–992, 2011.
- [110] J. A. List and C. A. Gallet, “What experimental protocol influence disparities between actual and hypothetical stated values?” *Environmental and Resource Economics*, vol. 20, no. 3, pp. 241–254, 2001.
- [111] J. J. Murphy, T. Stevens, and D. Weatherhead, “Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism?” *Environmental and Resource Economics*, vol. 30, no. 3, pp. 327–343, 2005.
- [112] J. L. Lusk and F. B. Norwood, “Bridging the gap between laboratory experiments and naturally occurring markets: An inferred valuation method,” *Journal of Environmental Economics and Management*, vol. 58, no. 2, pp. 236–250, 2009.
- [113] J. L. Lusk and F. B. Norwood, “An inferred valuation method,” *Land Economics*, vol. 85, no. 3, pp. 500–514, 2009.
- [114] D. A. Hensher, “Hypothetical bias, choice experiments and willingness to pay,” *Transportation Research Part B: Methodological*, vol. 44, no. 6, pp. 735–752, 2010.
- [115] M. Grieg-Gran, I. Porras, and S. Wunder, “How can market mechanisms for forest environmental services help the poor? preliminary lessons from latin america,” *World development*, vol. 33, no. 9, pp. 1511–1527, 2005.

- [116] R. Claassen, A. Cattaneo, and R. Johansson, “Cost-effective design of agri-environmental payment programs: US experience in theory and practice,” *Ecological Economics*, vol. 65, no. 4, pp. 737–752, 2008.
- [117] N. Kosoy, M. Martinez-Tuna, R. Muradian, and J. Martinez-Alier, “Payments for environmental services in watersheds: insights from a comparative study of three cases in Central America,” *Ecological Economics*, vol. 61, no. 2, pp. 446–455, 2007.
- [118] T. Kroeger and F. Casey, “An assessment of market-based approaches to providing ecosystem services on agricultural lands,” *Ecological Economics*, vol. 64, no. 2, pp. 321–332, 2007.
- [119] E. Gómez-Baggethun, R. De Groot, P. L. Lomas, and C. Montes, “The history of ecosystem services in economic theory and practice: from early notions to markets and payment schemes,” *Ecological Economics*, vol. 69, no. 6, pp. 1209–1218, 2010.
- [120] D. Aadland and A. J. Caplan, “Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias,” *American Journal of Agricultural Economics*, vol. 85, no. 2, pp. 492–502, 2003.
- [121] K. Brooks and J. L. Lusk, “Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data,” *American Journal of Agricultural Economics*, vol. 92, no. 4, pp. 1229–1241, 2010.
- [122] P. A. Champ and R. C. Bishop, “Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias,” *Environmental and Resource Economics*, vol. 19, no. 4, pp. 383–402, 2001.
- [123] D. Campbell, D. A. Hensher, and R. Scarpa, “Non-attendance to attributes in environmental choice analysis: a latent class specification,” *Journal of environmental planning and management*, vol. 54, no. 8, pp. 1061–1076, 2011.
- [124] R. Scarpa, T. J. Gilbride, D. Campbell, and D. A. Hensher, “Modelling attribute non-attendance in choice experiments for rural landscape valuation,” *European review of agricultural economics*, vol. 36, no. 2, pp. 151–174, 2009.
- [125] D. A. Hensher, A. T. Collins, and W. H. Greene, “Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding,” *Transportation*, pp. 1–18, 2012.
- [126] D. A. Hensher and W. H. Greene, “Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification,” *Empirical economics*, vol. 39, no. 2, pp. 413–426, 2010.

- [127] S. Hess, A. Stathopoulos, and A. Daly, “Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies,” *Transportation*, vol. 39, no. 3, pp. 565–591, 2012.
- [128] A. Poole. “The birds of North America online.” August 2005. [Online]. Available: <http://bna.birds.cornell.edu/BNA/>
- [129] K. E. Train, “EM algorithms for nonparametric estimation of mixing distributions,” *Journal of Choice Modelling*, vol. 1, no. 1, pp. 40–69, 2008.
- [130] S. Hess, M. Ben-Akiva, D. Gopinath, and J. Walker, “Advantages of latent class over continuous mixture of logit models,” working paper, 2011. URL: http://www.stephanehess.me.uk/papers/Hess_Ben-Akiva_Gopinath_Walker_May_2011.pdf, Tech. Rep., 2011.
- [131] C. R. Bhat, “An endogenous segmentation mode choice model with an application to intercity travel,” *Transportation Science*, vol. 31, no. 1, pp. 34–48, 1997.
- [132] W. H. Greene and D. A. Hensher, “A latent class model for discrete choice analysis: contrasts with mixed logit,” *Transportation Research Part B: Methodological*, vol. 37, no. 8, pp. 681–698, 2003.
- [133] J. Shen, “Latent class model or mixed logit model? A comparison by transport mode choice data,” *Applied Economics*, vol. 41, no. 22, pp. 2915–2924, 2009.
- [134] D. Pacifico *et al.*, “lologit: A Stata module for estimating latent class conditional logit models via the Expectation-Maximization algorithm,” *The Stata Journal*, *Forthcoming*, 2012.
- [135] W. M. Hanemann, “Welfare evaluations in contingent valuation experiments with discrete responses,” *American journal of agricultural economics*, vol. 66, no. 3, pp. 332–341, 1984.
- [136] S. K. Swallow, C. M. Anderson, and E. Uchida, “The Bobolink project: Selling public goods from ecosystem services using provision point mechanisms,” Zwick Center For Food and Resource Policy, Tech. Rep. 16, 2012.
- [137] J. M. Wooldridge, “Selection corrections for panel data models under conditional mean independence assumptions,” *Journal of Econometrics*, vol. 68, no. 1, pp. 115–132, 1995.
- [138] H. Uchida, C. A. Roheim, H. Wakamatsu, and C. M. Anderson, “Do Japanese consumers care about sustainable fisheries? evidence from an auction of eco-labelled seafood,” *Australian Journal of Agricultural and Resource Economics*, 2013.

- [139] D. Rondeau, G. L. Poe, and W. D. Schulze, “VCM or PPM? a comparison of the performance of two voluntary public goods mechanisms,” *Journal of Public Economics*, vol. 89, no. 8, pp. 1581–1592, 2005.
- [140] R. T. Carson and T. Groves, “Incentive and informational properties of preference questions,” *Environmental and Resource Economics*, vol. 37, no. 1, pp. 181–210, 2007.
- [141] C. A. Vossler and S. B. Watson, “Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field,” *Journal of Economic Behavior & Organization*, vol. 86, pp. 137–147, 2013.
- [142] A. Gelman, X.-L. Meng, and H. Stern, “Posterior predictive assessment of model fitness via realized discrepancies,” *Statistica Sinica*, vol. 6, pp. 733–807, 1996.