THREE ESSAYS ON MODELING

CONSUMER DEMAND

BY

JAE BONG CHANG

Bachelor of Science
SungKyunKwan University
Seoul, Korea
1999

Master of Science
Seoul National University
Seoul, Korea
2001

Master of Science
Iowa State University
Ames, Iowa
2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2009

THREE ESSAYS ON MODELING

CONSUMER DEMAND

Dissertation Approved:

Dr. Jayson L. Lusk

Dissertation Advisor

Dr. B. Wade Brorsen

Dr. F. Bailey Norwood

Dr. Lee C. Adkins

Dr. A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGMENTS

Finally, I face the most enjoyable moment of the long voyage, “writing a dissertation.” Many people played a role in the completion of this voyage, directly and indirectly. First, there is “the Captain” who guided and taught me the skills that enabled me to survive my long journey: I would especially like to gratefully and sincerely acknowledge my advisor, Dr. Jayson L. Lusk. He taught me to not only how to sail and navigate a boat but also how to develop techniques for confronting unexpected practical problems. During my voyage, the sea was not always gentle and the sky was not always bright and sunny. Sometimes, I had experienced strong winds and bad weather on the open sea. However, his mentorship was paramount in guiding in getting ability to tackle such challenges and providing a well rounded experience consistent my long-term career goals, “Be a captain of my own boat.” For everything you have done for me, Dr. Lusk, I thank you.

I would also like to thank all of the members of my committee. Whenever the fog has been dense for days, there is “the lighthouse”, Dr. B. Wade Brorsen, who showed the way in such thick fog and the light from him was charted to have a long and wide range. Dr. F. Bailey Norwood and Dr. Lee C. Adkins who went before me and still voyage kept me company through this journey provided not only essential practical advice but also the inspiration to pull through. Additionally, I would like to thank Dr. Oh Sang Kwon who introduced me to sailing in the “Academia.”

I would like to thank the Department of Agricultural Economics at Oklahoma State University for giving me the opportunity to board a ship and serve as a "crew." I am very grateful for the friendships I made along the way, especially Phumsith and Karina, who provided for some needed humor and entertainment in what could have otherwise been a somewhat stressful sailing environment.

Finally, and most importantly, I would like to thank my family back in South Korea. When I was blown off course and had been drifted, my late father, Byung Joo Chang, helped me to find my route. He was my Polaris and always with me at any time in the sky. My mom, Bok Ja Choi, always prays to God for my safe journey. Her support, encouragement, quiet patience, and unwavering love make me possible to return to port. I am very thankful for my sisters and brother-in-laws, Ji Eun, Cheorl-Ho Kim, Yoon Jung, Jin Ho Seok, Yoon Jin, and Jae Hyung Ko who have continued to have faith in me and to support and encourage me. Also, I thank my lovely nephews and nieces, Dong Hyun, Min Soo, Ji Won, Sung Won, Jun Hwa, Jun Suh, for their unending love.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Modeling consumer demand is important for businesses and public policy alike and has thus occupied much academic attention. The traditional neoclassical consumer theory assumes consumers' utility is a function of the quantity of goods consumed. Recent research has focused instead on the utility consumers derive from quality attributes of products, and posits that consumers make a discrete choice of one of a finite set of discrete bundle of attributes. This new framework poses challenges to develop demand specifications based on discrete choice modeling.

Since developed by McFadden (1973), the random utility model (RUM) has been the dominant theoretical framework for studying consumer behavior in discrete choice contexts (Louviere, Hensher, and Swait 2000). While the RUM assumes, as the standard neoclassical consumer model, that the decision maker acts rationally, it departs from the neoclassical theory by: (a) modeling a single choice among a finite set of mutually exclusive alternatives, and (b) incorporating uncertainty such as unobserved attributes of individuals (Bockstael and McConnell 2007). The RUM represents the integration of consumer behavior and randomness, and specifyies utility in terms of deterministic and random influences, making it ideally suited for the econometric analysis of choice.

Over the past decades, a great deal of attention has been devoted to refining consumer preference elicitation methods. An individual may choose her preferred alternative, rate how likely she would prefer the alternative on a cardinal scale, or rank the alternatives with or without monetary payment. In addition, there have been extensive developments in discrete choice econometrics to determine the level of utility of each alternative. Although the multinomial logit model has been known as a standard discrete choice model, researchers have made the effort to provide behavioral realism and to determine the empirical validity of choice models. As results, a number of competing discrete choice models have emerged. However, there remain many unanswered questions related to the use of preference elicitation methods and associated econometric models widely used to model consumer choice.

This study seeks to answer to some of unsolved issues in preference elicitation methods and discrete choice econometrics, and to provide a better understanding of consumer demand behavior. This dissertation consists of three essays. The first essay examines the external validity of survey and experimental methods to elicit consumer preference. Economists are often skeptical of stated preference methods. Although such methods have gained wide-spread popularity, there are relatively few studies examining the validity of these methods. This study compares the ability of the following methods to predict the market share of new and preexisting products in a grocery store: (a) hypothetical choice experiments, (b) non-hypothetical choice experiments, and (c) non-hypothetical ranking experiments. Moreover, for each of these elicitation methods, this study compares the predictive performance of the three econometric models: the multinomial logit, the independent availability logit, and the random parameter logit models.

The second essay departs from the traditional demand paradigm and seeks to determine whether consumer demand is a function of concerns for “fairness.” The essay investigates whether and how consumers prefer the distribution of benefits be shared across the participants in the food marketing channel. Although experimental studies have reported a wide array of other-regarding behavior, the pervasiveness of such behavior in the field is an open question. Using data from a mail survey, people’s preferences, when purchasing food products, for the distribution of benefits accruing to participants in the food supply chain – farmers, agribusiness processors, supermarkets, and consumers - are estimated. This study compares the ability of several inequality aversion models proposed in the general economics literature to explain food choice, and offers extensions to these models to better fit the food context. Moreover, we elicit consumers’ perceptions about the distribution of benefits resulting from the sales of non-organic and organic food and determine the extent to which preferences for the distribution of benefits can explain preferences for organic food.

The third essay of this dissertation explores the reliability of the random parameter (or mixed) logit model. Random parameter logit models are increasingly being reported in the literature and have become the norm in modeling choice data. This transition has happened primarily as a result of the conceptual advantages of the random parameter logit over the multinomial logit, the fact that results from the random parameter logit models have straightforward interpretations as compared to some other models that relax independence of irrelevant alternatives assumption, and because the model is supported by variety of econometric software packages. Despite these conceptual advantages, most practitioners who have estimated a random parameter logit model are aware that the estimates can be sensitive to specification and that achieving convergence is not always easy, in part because the likelihood function is not necessarily globally concave. Using a Monte Carlo analysis, this study determines the sensitivity of random parameter logit estimates. In particular, this study examines (a) how accurate are RPL estimates relative to true parameter values when there is no, low, and high preference heterogeneity, and (b) how the accuracy of random parameter logit estimates varies with sample size, number of replications used in the simulated maximum likelihood function, and econometric software package.

Overall, the results of this dissertation should improve understanding of consumer choice behavior.

CHAPTER II

EXTERNAL VALIDITY OF HYPOTHETICAL SURVEYS AND LABORATORY EXPERIMENTS

Introduction

There is perhaps no more important question for researchers working with survey and experimental preference elicitation methods than whether the elicited values accurately predict real-world, field behavior. A great deal of attention has been devoted in recent years to refining preference elicitation methods, including developments in contingent valuation, conjoint analysis, choice experiments, and experimental auctions. Although a great deal has been learned, there are very few studies examining the external validity of these methods. As such, skepticism surrounding stated and experimental willingness-to-pay values abounds. Practitioners advocating such methods have not adequately established the validity of the methods. The primary purpose of this paper is to address these concerns in a specific context and determine whether results from hypothetical and real-money purchasing experiments accurately reflect shopping behavior in a grocery store, and in the process, we also address some key issues for practitioners of revealed and stated preferences.

Why might behavior in laboratory experiments differ from behavior in the field? Experiments often involve unfamiliar preference elicitation methods and may impose constraints on people that they wouldn't normally encounter in the field. This is to say that the context of the laboratory experiment often differs from the field in ways that may have a substantive influence on behavior. For example, in an experimental setting people know their behavior is being scrutinized and social concerns may lead people to give "socially acceptable" answers. As another example, experimental exercises may omit goods that factor prominently in consumers' decision making processes. Furthermore, there may be differences in the type of people who participate in laboratory experiments and those who participate in field markets. People self-select into field markets, but are often recruited to participate in surveys or experiments. A number of recent papers have provided detailed discussions on factors affecting the divergence in laboratory and field behavior, focusing on differences in the subject pool, sample selection, the nature of the decision task, information, the nature and extent of scrutiny, and social norms (Harrison and List 2004; Levitt and List 2007; List 2006). Such discussions suggest that we should not always expect behavior to be identical in the lab and field, and that certain economic models might explain differences in the two environments. Nevertheless, if there is no correspondence between lab and field behavior, we must question what it is that is really being measured in the laboratory and whether it has any relevance for the "real world." Furthermore, many people conduct surveys or value elicitation experiments in the laboratory for the sole purpose of making predictions about what people will do in the field. As such, it is prudent to ask whether behavior in surveys/experiments accurately reflects behavior in the field, fully acknowledging that we cannot control for every difference between the two settings.

One of the key empirical findings that has bolstered criticisms of survey-based methods is that of hypothetical bias; the finding that willingness-to-pay elicited from hypothetical decision tasks almost always exceeds willingness-to-pay elicited from non-hypothetical decision tasks (e.g., see the reviews in Little and Berrens 2004; List and Gallet 2001; and Murphy et al. 2005). One interpretation of these findings is that only those values that can be elicited in non-hypothetical settings such as experimental markets are valid. The implicit assumption is that non-hypothetical willingness-to-pay is the "true" value that would correspond with actual payments in the marketplace or votes at the poll. However, neither hypothetical nor non-hypothetical valuation approaches are without their flaws and it is far from clear which approach, if either, is reflective of people's real world shopping behavior.

In addition to the explosion of preference elicitation methods in recent years, there has been a parallel development in the econometrics of discrete-choice models. Since the work of McFadden (1973), the standard in discrete choice modeling has been the multinomial logit (MNL). For years, however, people have questioned some of the restrictive assumptions of the MNL, which has led to a variety of competing models, almost all of which are generalizations of the MNL. One such example is the random parameter (or mixed) logit (RPL), which relaxes the assumption of the independence of irrelevant alternatives by modeling preference heterogeneity (McFadden and Train 2000). Another example is the choice-set consideration or the independent availability logit

(IAL), which relaxes the assumption of a deterministic choice set (Andrews and Srinivasan 1995; Haab and Hicks 1997; Swait and Ben-Akiva 1987). Although the RPL and IAL have been found to exhibit superior in-sample fit compared to the MNL (e.g., Revelt and Train 1998; Swait and Ben-Akiva 1987), better in-sample fit need not imply better out-of-sample predictive performance. The MNL is a more parsimonious model than the RPL or the IAL, and more parsimonious models are often found to exhibit better out-of-sample predictive performance (e.g., Kastens and Brester 1996; Murphy, Norwood, and Wohlgenant 2004). This suggests the need to investigate the ability of the MNL to predict field behavior as compared to more flexible model specifications.

In this paper, we compare the ability of the following methods to predict the market share of new and pre-existing products in a grocery store: (a) hypothetical choice experiments of the type frequently employed in survey work, (b) non-hypothetical choice experiments of the type frequently employed in laboratory experiments, and (c) a new non-hypothetical ranking experiment introduced by Lusk, Fields, and Prevatt (2008).[1] For each of these elicitation methods, we also compare the predictive performance of three econometric models: the multinomial logit (MNL), the independent availability logit (IAL), and the random parameter logit (RPL) models. We find that data collected from the non-hypothetical ranking method and analyzed via the MNL or RPL yield the best forecasts of retail market shares as indicated by mean-squared error and out-of-sample log-likelihood function values. Overall, results suggest a high level of external validity for certain methods and models, a finding which should increase confidence in economists' abilities to model market behavior with survey and experimental data.

Background

A number of studies have investigated the external validity of survey and experimental methods. A few studies have compared hypothetical contingent valuation survey responses to real, field public referenda (e.g., Johnston 2006; Vossler et al. 2003; Vossler and Kerkvliet 2003). These studies have found that contingent valuation responses map reasonably well to observed voting behavior depending on how one chooses to model and code "indifferent" and "don't know" responses with the contingent valuation method.[2] Although these studies are useful in their own right, the findings may not be particularly relevant to valuing private attributes related to new products, technologies, and food policies. The incentives for people to give truthful and accurate answers can differ markedly as one moves from public to private goods and as one moves from referenda-type questions to choice-experiment-type questions (Carson and Groves 2007).

Of more direct relevance to the current investigation are the few previous studies that have compared non-hypothetical experimental behavior to real, field shopping behavior (e.g., see Brookshire, Coursey, and Schulze 1987; Lusk, Pruitt, and Norwood 2006; Shogren et al. 1999). Brookshire, Coursey, and Schulze (1987) compared demand curves constructed from bids for strawberries collected in a laboratory auction to implied demand curves from actual purchases of strawberries made via door-to-door sales. They were unable to reject the hypothesis that the valuations from the auction were no different than the field sales data. Their findings thus implied that valuations were stable across setting (the lab versus the field) and elicitation method (auction versus purchases at a stated price).

Shogren et al. (1999) compared data from a mail survey and a non-hypothetical lab valuation exercise to grocery store purchases of irradiated chicken. They found higher levels of acceptability of irradiated chicken in both the survey and experimental market as compared to the retail setting when irradiated chicken was sold at an equal or discounted price relative to non-irradiated chicken. Choices in the hypothetical survey and non-hypothetical lab experiment were more similar to grocery store behavior when irradiated chicken was sold at a premium over regular chicken. More precisely, Shogren et al. (1999) found that 80% of survey and experiment participants preferred irradiated to non-irradiated chicken breasts when the two products were offered at the same price; however, only about 45% of shoppers in the retail setting bought irradiated chicken when it was priced the same as non-irradiated. In contrast, in all three settings (survey, experimental market, and store) about 33% of people bought the irradiated chicken when it was priced at a 10% premium over regular chicken.

Results from Shogren et al. (1999) suggest mixed findings regarding the external validity of both hypothetical and non-hypothetical responses and their results seem to suggest that hypothetical CV responses and non-hypothetical experiments performed about equally well in predicting retail market share. However, formal statistical tests were not carried out to determine whether one method outperformed the other. Further, their finding of close similarity of real and hypothetical responses stands in stark contrast to the typical finding on this issue (e.g., see Fox et al. 1998 for evidence on hypothetical bias for irradiated food). Additionally, and most importantly, the fact that experimental and survey participants received information about irradiation created a confound in the comparison of survey/lab behavior and field behavior as the subjects in the survey/lab were informed whereas those in the grocery store were generally not.

This study builds on previous literature in a number of ways. First, we utilize what is becoming a standard valuation approach: the choice experiment. Second, we carry out formal statistical tests to determine whether real or hypothetical responses to a choice experiment better predict actual retail sales. Third, unlike most previous studies that focus on a single good, we compare predicted and actual market share for twelve goods in three distinct product categories including three new goods previously unavailable for sale in the local market. Fourth, we explicitly refrain from providing experimental subjects any more information about the products than what typical shoppers would have in the grocery store. Fifth, in addition to the choice-experiment approaches, we investigate the external validity of a new non-hypothetical conjoint ranking mechanism. Finally, this study considers the out-of-sample performance of several competing econometric models.

Methods and Procedures

Hypothetical and Non-Hypothetical Valuation Exercises

As previously mentioned, one of the key factors that could cause a divergence between behavior in laboratory experiments and behavior in the field is differences in the people in the two environments. In an attempt to ensure that the people in our experiment were generally reflective of shoppers in the grocery store serving as our field referent, we did not rely on a student sample. A random sample of people from Stillwater, Oklahoma was recruited to participate in the hypothetical and non-hypothetical valuation exercises. The Bureau for Social Research at Oklahoma State University used random digit dialing techniques to contact people and request their participation in a "food preference study" in exchange for $40. Approximately 35% of the people contacted agreed to participate in the study. Recruited participants were mailed a reminder note and a map to the study's location. Upon arrival at the study site, subjects were randomly assigned to one of three experimental treatments: hypothetical choice, non-hypothetical choice, or non-hypothetical ranking.[3] In total, 47 consumers participated in the hypothetical choice treatment, 46 people participated in the non-hypothetical choice treatment, and another 42 subjects were assigned to the non-hypothetical ranking treatment.

The demographic characteristics of the participants, in terms of gender, education, income, race, and presence of children in the household were similar across the three treatments; we could not reject the null hypothesis of equality of these characteristics across the three experimental treatments. In addition, the characteristics of participants in our experiment are similar to those of shoppers in the store. To characterize the characteristics of store shoppers, we utilize data collected by Lusk, Norwood, and Pruitt (2006) from over 440 randomly intercepted shoppers in the store. The average age in the store was 46.6 and the average age of our experimental participant was 46. Half the store shoppers had an income less than $40,000 per year, and 59% of our experimental participants were in the same category. The fraction of the sample that was white was 87% in both the store and the experiment. Twelve percent of store shoppers were college students; thirteen percent of our experimental sample was students. Thus, our sample is reasonably representative of shoppers in the local grocery store, certainly more so than would have been the case had our experiment used a convenience sample of students. This does not mean that no students participated in our experiment, only that they participated proportionately to their share of the local population.

Regardless of the treatment to which an individual was assigned, they were requested to investigate 12 products that were located in the front of the room. Other than what could be ascertained from the packages, subjects were not given any additional information about the products. The 12 products were grouped into three product categories: dishwashing liquid, ground beef, and wheat flour. In each product category, there were three pre-existing brands and one new brand that was not available for sale in the local market. The three new products were: (a) Eco-Plus, an "environmentally friendly" dishwashing liquid, (b) Cattle Tracks, an organic ground beef brand, and (c) GO Organic wheat flour, an organic and regionally grown brand.

The products used in the experiment were chosen to mirror those offered in a local grocery that agreed to participate in the research. The local store sold each of the 12 goods, and allowed us to control the goods' prices and obtain information on sales volume. In the ground beef product category, the store sold only three types and we utilized all three in our experiments (fresh, lean, and diet lean) in addition to the new organic brand. The same was true of the whole wheat flour category: there were three pre-existing brands for sale and we utilized these in our experiment to mimic what was available in the grocery store (the brands were: Gold Medal, Hodgson Mill, and King Arthur). In the dishwashing liquid category, there were over 30 competing products on the store shelf. From these, we selected the highest selling product from each of the main brand names (Dawn, Joy, and Palmolive).

The set-up of the hypothetical and non-hypothetical choice treatments was similar to that in Lusk and Schroeder (2004). In both treatments, subjects were asked to answer five choice questions for each product category (dishwashing liquid, ground beef, and wheat flour). Thus, in total each participant answered 15 discrete choice questions (or shopping scenarios) regarding which product they wanted at the set of prices in the respective choice set. In each choice set were five options including a "none" or "no purchase" option. Prices of each of the goods were varied between \$2, \$3, and \$4, which encompassed the range of prices for these products in the grocery store. An orthogonal fractional factorial design was used to assign prices to products, ensuring that the prices each of the products were uncorrelated with each other across the design. To achieve a perfectly orthogonal design, nine choice options had to be used. We systematically varied these nine options across surveys so that each person only answered five choice questions, but the complete design was repeated several times across subjects in each treatment. Figure 1 shows an example of a discrete choice question for each of the three product categories. The order in which the brands were presented in the choice tasks was varied across participants so that our results were not unduly influenced by a possible order effect. In the hypothetical choice treatments, subjects were told:

> In each scenario, you should choose ONE of the products you would like to purchase or you can choose not to purchase any of the products by checking the last option in each shopping scenario. For each scenario, assume that you have the opportunity, here and now, to purchase ONE and ONLY ONE of the items at the listed price. While you will not actually buy any products today or pay the posted prices, please respond to each shopping scenario as if it were a real purchasing opportunity and you would have to give up real money were one of the 15 scenarios to be selected as binding.

In the non-hypothetical choice treatments, subjects were told:

> After everyone completes all 15 shopping scenarios, we will ask for a volunteer to draw a number (1 to 15) from a hat to determine which shopping scenario will be binding. In the hat are numbers 1 through 15. If the number 1 is drawn then the first shopping scenario will be binding. If the number 2 is drawn the second shopping scenario will be binding, and so on. For the binding scenario, we will look at the product you have chosen, give you your chosen product, and you will pay the listed price in that scenario. If you choose "none" you will not receive a product and you will pay nothing.
>
> Note: This is a real decision making exercise. For the randomly selected shopping scenario, we will really give you the chosen product and we really expect you to pay the price. The price will be deducted from your $40 participation fee. Although only one of the 15 shopping scenarios will be binding there is an equal chance of any shopping scenario being selected as binding, so think about each answer carefully.

Although non-hypothetical choice experiments are incentive compatible, they do not provide a great deal of information. That is, the researcher only observes the single most preferred option in a choice experiment. By contrast, a ranking-based application contains a consumer's complete preference ordering. Thus, a ranking approach might be more accurate than a choice-based approach because it contains more information. However, a ranking task differs from what people usually do at the grocery store. Thus, an empirical comparison of which method (choice versus ranking) is important because both approaches have strengths and weaknesses and neither can be ruled out *a priori* as inferior.

The third treatment utilized the non-hypothetical conjoint ranking approach introduced by Lusk, Fields, and Prevatt (2008). In the non-hypothetical ranking treatment, subjects were similarly asked to respond to 15 shopping scenarios. However, instead of indicating which one product they most desired in each scenario, people were asked to rank the products in terms of the relative desirability. People were asked, for each choice set, to put a 1 next to the product which they most preferred, a 2 next to the second most preferred product, and so on. The non-hypothetical ranking valuation questions were similar in appearance to those shown in figure 1 with the exception that people were asked to rank the options instead of choosing one option.

To ensure that the ranking task was incentive compatible, one of the 15 ranking scenarios was selected as binding by drawing a number from a hat. Then, for the binding scenario, a second number was randomly drawn to determine which product the participant purchased. In particular, after determining which of the 15 scenarios was binding, a second random number (1 through 36) was drawn. If one of the numbers 1-15 was drawn, the participant purchased the product they ranked first, if one of the numbers 16-25 was drawn, they purchased the product they ranked second, if one of the numbers 26-30 was drawn, they purchased the product ranked third, if the one of numbers 31-34 was drawn, they purchased the product ranked forth, and if 35 or 36 was drawn, they purchased the product ranked last. In this way, there was a higher chance a participant purchased a product they gave a higher rank (e.g., the chance of purchasing a product ranked 1st is higher than the chance of purchasing a product ranked 2nd and so on). It is easy to see that a person is always better off ranking their most preferred item first, because the chances of it being chosen is the highest, ranking the second most preferred product second, and so on. The least preferred item should be ranked last because it has the lowest chance of being selected.

Retail Market

We obtained agreement from a local grocery store in Stillwater, Oklahoma to participate in the research. Approximately two weeks after the laboratory experiments, store managers introduced the three new items (the environmentally friendly dishwashing liquid, organic ground beef, and organic whole wheat flour) and gave each a prominent shelf position. The grocery store did not sell any other organic or environmentally friendly products in these product categories. Each of the products was placed on sale for $4.00.

The store kept the new products on the store shelves for one month. During this period, we requested that the store hold constant the prices of the new and pre-existing products in each product category. The authors visited the store each day that the products were sold to record the prices and ensure that each product was stocked. After a month-long time period, the store provided us with sales data from the three product categories (fresh ground beef, dishwashing liquid, and flour) aggregated over the month time period. With these data, we are able to calculate the (quantity) market share of each good in each product category. Because the store failed to completely hold constant the prices of all pre-existing products during the time period, we used the store's scanner data to calculate the weighted average price of each good over this time period. These actual purchase shares can be directly compared with the predicted shares resulting from the laboratory experiments.

Econometric Models

Based on the random utility model, the *i*th consumer's utility of choosing option *j* is

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{2.1}$$

where V_{ij} is a deterministic component and ε_{ij} is an iid stochastic component. In this application, the deterministic portion of the utility function can be expressed as

(2.2) $$V_{ij} = \alpha_j + \alpha_{price} P_{ij}$$

where α_j is an alternative specific constant indicating utility for alternative *j* relative to an omitted option, α_{price} represents the marginal utility of price, and P_{ij} is the price of alternative *j* for consumer *i*.

Multinomial Logit Model

Assuming that the ε_{ij} are distributed Type I Extreme Value yields the familiar MNL, where the probability of consumer *i* choosing option *j* out of a total of *J* options is:

(2.3) $$\text{Prob}\{j \text{ is chosen}\} = \frac{\exp(V_{ij})}{\sum_{k=1}^{J} \exp(V_{ik})}.$$

Equation (2.3) is the appropriate model in the treatments where people made a discrete choice between options, however, the third treatment involved individuals ranking alternatives. The ranking data can be easily analyzed in this framework using the rank-ordered logit model, which is a straightforward extension of the MNL. In particular, Beggs, Cardell, and Hausman (1981) show that out of a set of *J* options, the probability that option 1 is preferred to option 2, option 2 is preferred to option 3, option 3 is preferred to option 4, and so on is given by:

(2.4) $$\prod_{j=1}^{J-1} \frac{e^{V_{ij}}}{\sum_{k=j}^{J} e^{V_{ik}}}$$

which is simply the product of *J*-1 multinomial logit models.

Independent Availability Logit Model

The MNL assumes deterministic choice sets, meaning that it is assumed that all consumers consider all options presented to them. However, some people may only consider a subset of all available options. If so, the MNL formula in (2.3) will be incorrect because the choice probabilities are calculated by summing over the utility of all *J* goods.

To model this behavior, a probabilistic model for the choice set generation process can be formulated following Manski (1977). The formulation distinguishes between the choice set presented in the research instrument and the *consideration* set, the latter of which contains a subset of all available options encompassing all the items people might actually consider. An individual's true consideration set cannot be known with certainty, but their choice behavior can be used to make probability statements about the likelihood of competing consideration sets being the true choice set. Manski (1977) details such an estimator, with applications that can be found in Swait and Ben-Akiva (1987) and Ben-Akiva and Boccara (1995). Louviere, Hensher, and Swait (2000) show that the probability of individual *i* choosing option *j* in the IAL model is

$$\text{(2.5)}\qquad \text{Prob}\{j\,\text{is chosen}\} = \sum_{C_i \subseteq C} \text{Prob}(j|C_i) \times \text{Prob}(C_i) = \frac{\exp(V_{ij})}{\sum_{k \in C_i} \exp(V_{ik})} \times \frac{\prod_{j \in C_i} A_{ij} \prod_{j \in C - C_i} (1 - A_{ij})}{1 - \prod_{j \in C} (1 - A_{ij})}$$

where C is the set of all deterministically feasible consideration choice sets, C_i is consumer *i*'s true consideration set, and A_{ij} is the probability that alternative *j* is available and present in the true choice set for the consumer *i*. Equation (2.5) shows that the probability of choosing an option is determined by calculating the probability of all possible consideration sets being the true choice set (in our application there are $(2^5 - 1) = 31$ possibilities) and for each possibility, calculating the probability that the alternative is chosen. We parameterize A_{ij} as follows

$$A_{ij} = \frac{1}{(1+\exp(-\beta_j))} \tag{2.6}$$

where β_j is an alternative-specific constant. Although the IAL relaxes the assumption of deterministic choice sets, it assumes that the presence/absence of one alternative in the choice set is independent of the presence/absence of another alternative.

To implement the IAL with the ranking data, we followed studies such as Boyle et al. (2001) and "explode" the ranking data by converting the ranks into choices. For example, for the product ranked first, it was assumed this product would be chosen out of all five alternatives. For the product ranked second, it was assumed that it would have been chosen as most preferred out of the remaining four options. For the product ranked third, it was assumed it would have been chosen as most preferred out of the remaining three options. Finally, the product ranked fourth was assumed to have been chosen from the remaining pair of options. Thus, each ranking is "exploded" into four choices, which are then used to estimate the IAL.[4]

Random Parameter Logit Model

The MNL assumes preference homogeneity in the sample, implying that all coefficients of the utility expression in equation (2.1) are the same across individuals. The IAL allows for some heterogeneity in the extent to which people differ in terms of the alternatives they consider. The random parameters logit (RPL) model allows a more flexible and continuous form of preference heterogeneity, where utility coefficients vary across individuals according to continuous probability distribution functions.

The RPL is implemented by specifying the alternative-specific constants shown in equation (2.2) as

$$\alpha_{ij} = \overline{\alpha}_j + \sigma_j v_{ij} \tag{2.7}$$

where $\overline{\alpha}_j$ is the population mean alternative specific constant for option j, σ_j is the standard deviation of the distribution of the coefficient α_{ij} around the population mean, and v_{ij} is a stochastic term which is distributed normally with zero mean and standard deviation one. As in Revelt and Train (1998), we assume the price coefficients are invariant across individuals. As shown by Train (2003), the probability of choosing option j is

$$\text{Prob}\{j \text{ is chosen}\} = \int \frac{\exp(V_{ij})}{\sum_{k \in C_i} \exp(V_{ik})} f(\alpha_i) d\alpha_i \tag{2.8}$$

where $f(\alpha_i)$ is the density of the coefficients α_i. Because equation (2.8) lacks a closed form solution, the parameters of the model are estimated by simulated maximum likelihood estimation techniques following Train (2003). As with the IAL model, to estimate the parameters of the RPL on the ranking data, we utilize the "exploded" ranking data converted into choices.

Comparing Experimental Behavior to Retail Market

The MNL, IAL, and RPL models can be used to calculate the predicted market share for each product based on equations (2.3), (2.5), and (2.8). Once the parameter estimates from these models are obtained, the predicted share can be estimated by substituting these coefficients into probability equations, given the prices utilized in the store.

Calculating the true, field market share from the grocery store is straightforward. Sales data provided by the local grocery store contain the total volume and weighted-average price of each good in each product category sold. The total sales volume figures were used to calculate the quantity share each product received in each product category by simply dividing the sales of each good by total sales in the product category. The weighted average prices of Dawn, Joy, Palmolive, and Eco-Plus in the store were \$1.99, \$1.99, \$2.89, and \$4, respectively. Fresh, Lean, Diet Lean, and Organic ground beef were sold at prices of \$1.76, \$2.16, \$2.58, and \$4 per pound, respectively. The prices of Hodgson Mill, King Arthur, Gold Medal, and GO Organic wheat flour were \$2.99, \$3.99, \$2.65, and \$4, respectively.

To evaluate which elicitation method most closely predicted the real market shares, two criteria were used. First, we calculated the mean squared error (MSE), which is simply the mean of the squared difference between the predicted and actual shares in each product category. The elicitation method and econometric model with the lowest MSE is deemed to have the best predictive performance. In addition to this criterion, we also utilized the out-of-sample log likelihood function (OSLLF) approach (Norwood, Lusk, and Brorsen, 2004). The OSLLF criterion selects the models with the highest likelihood function values at out of sample observations. In this study, the OSLLF can be calculated as:

$$OSLLF = \sum_{j=1}^{J} TM_j \ln(EM_j) \tag{2.9}$$

where TM_j is the true market share from the grocery store and EM_j is the estimated market share for good j for a particular product category, elicitation method, and estimation method.

To test the hypothesis of whether the MSE or OSLLF differs across elicitation/estimation method, standard errors or 95% confidence intervals must be calculated. For the MNL and IAL models, 95% confidence intervals on the MSE and OSLLF are calculated via parametric bootstrapping following Krinsky and Robb (1986). Calculating such statistics for the RPL involves simulating a population of consumers for each parameter bootstrap.[5] In addition to the 95% confidence intervals, we make use of the combinatorial re-sampling approach described in Poe, Giraud, and Loomis (2005) by utilizing the bootstrapped values from the MNL, IAL, and RPL models to test the hypothesis that the MSE/OSLLF is lower or higher in one method versus another.

Results

Table II-1 reports estimates of nine MNL models (three elicitation methods × three product categories). For each elicitation method and product category, the price coefficient is negative, meaning higher prices are associated with a lower likelihood of purchase. The alternative specific constants are estimated to indicate the utility of each option relative to the "none of these" option. These parameters are generally positive, meaning that holding price constant, people preferred having one of the products to having nothing at all. The hypothesis that all parameters are zero is rejected by a likelihood ratio test (p-value < 0.01) for all nine models shown in table II-1.[6]

Table II-2 presents the estimates of the IAL and RPL models. For the IAL model, coefficients in the availability function are also estimated in addition to the alternative specific constants in utility function. Positive parameters in the availability function imply a higher likelihood of a particular alternative being in the consideration choice set. For example, availability coefficients for Palmolive, Diet Lean, and Gold Medal are negative, indicating that those products are less likely to be in the true choice set.

For the dishwashing liquid and ground beef categories, an interesting pattern of results emerges in the IAL. In particular, for hypothetical choices, the alternative-specific constants for the new products have negative signs in the utility function (i.e., $\alpha_4 < 0$), but positive signs in the availability function (i.e., $\beta_4 > 0$). However, for the non-hypothetical methods, the opposite is true (i.e., $\alpha_4 > 0$ and $\beta_4 < 0$). This means that in hypothetical situations, "none" is unlikely to be in the choice set (however, in the unlikely event that "none" enters the set, it is likely to be chosen). The opposite is true for non-hypothetical situations: "none" is likely to be in the choice set. One interesting result stemming from the non-hypothetical treatments is that the availability parameters on the new goods, Eco-Plus and Organic beef, are negative even though the utility coefficients are positive. We interpret this result to mean that consumers who strongly care about organic products and/or the environment comprise a small fraction of the population, but their satisfaction from purchasing those products is much higher than from buying "conventional" products. That is, there is a small percentage of people who include Eco-Plus and/or Organic beef in their choice set, and for these people these products are strongly preferred.

Table II-2 also reports results for the mean and standard deviation estimates for each option for the RPL model. Results reveal large and statistically significant standard deviations for all products in every treatment, except for dishwashing liquid and wheat flour in the non-hypothetical choice method, implying a significant amount of preference heterogeneity. That the magnitude of the standard deviation of preferences for King Arthur flour (α_2) in the non-hypothetical treatment is extremely large, is indicative of the fact that only one subject chose this option in this particular treatment.

Table II-3 reports the predicted market shares for each product by experimental treatment and econometric model. The last column in table II-3 reports the actual market shares from the grocery store. Generally, we find that the predicted market shares from the MNL and RPL models correspond well with the actual market shares. The exception to this statement is that the MNL under-predicted the success of the new organic flour in the grocery store. In addition, the IAL tended to make very precise predictions, forecasting high market shares for a single product. Although the IAL did a good job predicting *which* product would receive the highest market share, it tended to perform poorly in terms of predicting outcomes over the entire product category. Despite this, the experimental data often perform remarkably well in predicting actual sales data. For example, the market share estimates for all products from the MNL, non-hypothetical ranking treatment for dishwashing liquid never diverge from the true market share by more than three percentage points for any product.

Table II-4 contains the key comparisons among the methods. Shown in table II-4 are the MSE and OSLLF for each experimental treatment, estimation method, and product category. Focusing first on the MSE criteria, for which a lower value is preferred, we find that for the dishwashing liquid category and the MNL model, the MSE for the non-hypothetical choice method is lowest at 0.001. However, for ground beef and whole wheat flour, the MSE for the non-hypothetical ranking treatment is lowest at 0.007 and 0.105, respectively. For the IAL model, the MSE from the non-hypothetical ranking method is lowest for dishwashing liquid and ground beef at 0.041 and 0.239, respectively, and the MSE in the non-hypothetical choice method is lowest for wheat flour at 0.221. For the RPL model, the MSE is lowest for the hypothetical choice method for dishwashing liquids, but is the highest for the ground beef and wheat flour categories.

The second selection criterion is the OSLLF method which can be used to rank methods/models by likelihood function values observed at out-of-sample (grocery store) observations. A higher OSLLF value is preferred. In the MNL, the OSLLF values for the non-hypothetical choice method are highest, -1.006 and -1.139, respectively for dishwashing liquid and ground beef categories. For the whole wheat flour categories, the OSLLF value for the non-hypothetical ranking method has the highest value, -1.561. For the IAL model, the OSLLF values for the hypothetical choice are highest for dishwashing liquids. However, overall, we can see that the IAL has poor predictive performance relative to the MNL and RPL. For the RPL, OSLLF values for the real ranking method are highest for dishwashing liquid and whole wheat flour categories.

Overall, the findings in table II-4 suggest that the hypothetical choice method performs relatively poorly at predicting market shares. We come to this conclusion by restricting attention to just the MNL or RPL models, which dominate the IAL in terms of predictive performance. Second, we note that within a product category, one can always find a lower MSE for the non-hypothetical choice as compared to the hypothetical choice when selecting the lowest value across the MNL and RPL models. For example, for dishwashing liquid, ground beef, and whole wheat flour, the lowest MSE values across the MNL and RPL models are 0.001, 0.004, and 0.191, respectively for the non-hypothetical choice, where the comparable figures are 0.014, 0.04, and 0.251 for the hypothetical choice method. Thus, so long as one has the freedom to choose the best econometric model, we find that making the choice task non-hypothetical significantly improves out-of-sample forecasts. Carrying out the same calculation for the non-hypothetical ranking task reveals that the lowest MSE values across the MNL and RPL models are 0.002, 0.005, and 0.01 for dishwashing liquid, ground beef, and whole wheat flour, respectively. Thus, the non-hypothetical ranking method performs about the same as the non-hypothetical choice for dishwashing liquid and ground beef, but much better for the whole wheat flour category.

We further summarize the findings in two ways. First, the last three rows in table II-4 show the results aggregated across all three product categories. Results reveal that within any particular econometric model, the MSE is the lowest and the OSLLF the highest for the non-hypothetical ranking method. Test statistics derived from the combinatorial re-sampling method of Poe, Giraud, and Loomis (2005) indicate that, for each econometric model, the MSE for the non-hypothetical ranking method is significantly lower and the OSLLF is the significantly higher than the hypothetical choice method ($p < 0.05$). The only exception to this statement is that there is no significant difference across elicitation methods if one only looks at the IAL estimates and uses the OSLLF criteria. Results also reveal that, in aggregate, the RPL model out-predicts the MNL and the IAL regardless of elicitation method. Results from the combinatorial re-sampling test indicate that the RPL yields lower MSE and higher OSLLF at the $p<0.01$ level for the non-hypothetical ranking method when data are aggregated across product category. For the hypothetical method, however, the differences in MSE and OSLLF between RPL and MNL are not statistically significant. For the non-hypothetical methods, in particular, the difference in OSLLF estimated from the RPL and IAL is significant, indicating superiority of the RPL model in the non-hypothetical data. Of course, the last three rows of table II-4 ignore differences across product category.

The second approach we use to summarize the results is to carry out an analysis of variance analysis (ANOVA) using the data in table II-4 to test the hypothesis that the three key variables (product category, elicitation method, and econometric model) and their interactions affect MSE and OSLLF. Results shown in table II-5 reveal a very high R^2 indicating variation in these variables explain virtually all the variation in MSE and OSLLF. Results reveal that all variables and their interactions (except the elicitation method × econometric method interaction) significantly influence MSE. A similar result is true of OSLLF; however, the only significant interaction affecting OSLLF is the product category × econometric model interaction. The F-values associated with the econometric model and product category are the largest, indicating that we can be most confident these two variables influence out-of-sample prediction performance. Table II-5 confirms the notion that making a decision task non-hypothetical improves out-of-sample prediction performance. Results also reveal that the RPL or the MNL exhibits the best predictive performance depending on the product category and elicitation method.

Conclusions

An important question in experimental economics, stated preference methods, and contingent valuation is whether elicited values accurately correspond with consumer behavior in "real" markets. This paper considers how three elicitation methods (hypothetical choice, non-hypothetical choice, and non-hypothetical ranking) and three econometric models (the MNL, IAL, and RPL) fared in predicting actual grocery store sales for twelve products in three product categories (dishwashing liquid, ground beef, and whole wheat flour). Our findings confirm the implicit assumption made in much of the work on hypothetical bias: the non-hypothetical choices are a better approximation of "true" preferences than are hypothetical choices.

Recent years have witnessed a trend toward estimating discrete choice models that relax the assumptions of the traditional MNL. Results suggest that relaxing these assumptions can not only improve in-sample fit, but can improve out-of-sample predictions as well. However, this is not always true. For example, for two out of the three product categories investigated, we found that the RPL exhibited the best predictive performance with the MNL being best in the third case. However, the IAL, which relaxes the assumption of deterministic choice sets, never out-performed the MNL or RPL within a product category or elicitation method despite the warnings by some that this assumption of the MNL is overly restrictive (e.g., see Haab and Hicks, 1997; Swait and Ben-Akiva, 1987). At this point, it is difficult to determine exactly why the IAL performed so poorly in predicting out-of-sample field behavior. It may be that people form their consideration sets differently in experimental and field situations. Indeed, the experimental setting, by its nature, frames a very narrow and precise consideration set. Imposing the structure of the estimated consideration set on field choices (where there were many more options) appears to yield poor predictions. Apparently, it is better (in term of out-of-sample prediction performance) to ignore modeling the choice-set generation process altogether than it is to model it incorrectly with overly restrictive experimental choice tasks. That the RPL performed better than the MNL in two out of three cases is likely attributable to the fact that the RPL is a more flexible model and that RPL accounts for the panel or repeated nature of the choice data where the MNL did not.

A final note to those wary of survey and experimental methods is in order. As shown in table II-3, all elicitation methods considered in this paper exhibited a reasonably high level of external validity. Take for example, the predicted market share for dishwashing liquid from the non-hypothetical ranking task and the MNL model. Our predicted market shares versus the true market shares were 0.50 vs. 0.50 for Dawn, 0.34 vs. 0.37 for Joy, 0.13 vs. 0.14 for Palmolive, and for the new Eco-Plus product, the predicted value was 0.03 and the actual share was 0.00. These findings should come as a welcome relief to agribusinesses in need of research to formulate pricing and marketing strategies and to policy makers in need of non-market values to determine the costs and benefits of various food labeling, food safety, and food nutrition policies.

Notes

1. An in-depth discussion of hypothetical choice experiments is provided by Louviere, Hensher, and Swait (2000) and the method has been used in recent literature by Adamowicz et al. (1998) and Lusk, Roosen, and Fox (2003). Alfnes et al. (2006), Ding, Grewal, and Liechty (2005), and Lusk and Shroeder (2004) all have implemented non-hypothetical choice experiments.

2. Other studies have compared hypothetical contingent valuation responses to indirect valuation methods such as hedonic analysis, travel cost methods, or other revealed-preferences methods (e.g., Adamowicz, Louviere, and Williams 1994; Brookshire et al. 1982; Carson et al. 1996; Loomis, Creel, and Park 1991).

3. Note that we did not compare responses under all four possible treatments (hypothetical choice, real choice, hypothetical rank, real rank); however, we still have an un-confounded test of the effect of real payments (within the choice experiment method) and an un-confounded test of the effect of moving from choices to rankings (given real payments). In terms of the hypothetical bias question, we chose to focus on what is now a standard valuation method: the choice experiment. In addition to this question, however, we were also interested in whether another method might improve on choice experiments. Because Lusk, Fields, and Prevatt (2008) already showed that real and hypothetical rankings can differ, we chose to focus our comparison of competing methods on real choices and real rankings.

4. Although some might object to this re-coding, it is important to note that estimating the rank ordered logit in equation (2.4) yields the same result as estimating the MNL in

equation (2.3) on the “exploded” rank data. That is, the assumption made in exploding the rank data is exactly the same as that made in estimating the rank-ordered logit.

5. To calculate the 95% confidence intervals on market share and MSE/OSLLF for the RPL, the following steps were taken: 1) a sample of 1,000 mean parameter vectors associated with $\overline{\alpha}_j$ and σ_j was drawn from the original parameter vector and covariance matrix of the estimated model, 2) for each of the 1,000 draws, a sample of 1,000 simulated individuals was created by drawing values of v_{ij} for each alternative (i.e., there are 1,000,000 generated observations), 3) for each sample of 1,000 simulated individuals, the mean market share and associated MSE/OSLLF was calculated, and 4) the 95% confidence intervals were determined by identifying the 25th and 975th highest mean MSE/OSLLF values across the 1,000 mean parameter draws.

6. Caution should be taken in directly comparing coefficients across elicitation methods as it involves a comparison of both the utility parameters and scale (Swait and Louviere 1993).

Shopping Scenario 1: Please check the ONE item you prefer

	Palmolive: Original Scent (25 fl oz)	Dawn: Original Scent (25 fl oz)	Joy: Original Scent (25 fl oz)	Eco Plus (28 fl oz)	**NONE**
	$3.00 ↓	**$2.00** ↓	**$4.00** ↓	**$4.00** ↓	**$0.00** ↓
I Choose (check one)	☐	☐	☐	☐	☐

Shopping Scenario 6: Please check the ONE item you prefer

	Diet Lean Ground Beef	Fresh Ground Beef	Lean Ground Beef	Cattle Tracks Organic Ground Beef	**NONE**
	$2.00 ↓	**$4.00** ↓	**$3.00** ↓	**$4.00** ↓	**$0.00** ↓
I Choose (check one)	☐	☐	☐	☐	☐

Shopping Scenario 11: Please check the ONE item you prefer

	Gold Medal Whole Wheat Flour	Go Organic Whole Wheat Flour	King Arthur Whole Wheat Flour	Hodgson Mill Whole Wheat Flour	**NONE**
	$3.00 ↓	**$4.00** ↓	**$2.00** ↓	**$2.00** ↓	**$0.00** ↓
I Choose (check one)	☐	☐	☐	☐	☐

Figure II-1. Examples of questions for the hypothetical and non-hypothetical choices

Table II-1. Multinomial Logit Model Estimates by Product Category and Elicitation Method

Commodities	Elicitation Method: Hypothetical Choice	Non-hypothetical Choice	Non-hypothetical Ranking
Dishwashing Liquid			
α_1, Dawn	4.81*[a]	1.63*	3.09*
	(0.48)[b]	(0.82)	(0.13)
α_2, Joy	4.06*	1.38	2.71*
	(0.47)	(0.70)	(0.12)
α_3, Palmolive	3.76*	1.79*	2.74*
	(0.45)	(0.78)	(0.13)
α_4, Eco-Plus	4.16*	1.66	2.53*
	(0.45)	(0.85)	(0.12)
α_{price}	-1.45*	-1.66*	-1.13*
	(0.15)	(0.33)	(0.04)
Log likelihood	-294.29	-164.97	-884.43
Ground Beef			
α_1, Fresh	4.02*	1.15	1.94*
	(0.44)	(0.85)	(0.23)
α_2, Lean	5.15*	2.76*	2.43*
	(0.45)	(0.65)	(0.24)
α_3, Diet Lean	3.99*	2.79*	2.02*
	(0.44)	(0.75)	(0.24)
α_4, Organic	3.64*	2.47*	1.77*
	(0.44)	(0.77)	(0.23)
α_{price}	-1.28*	-1.85*	-0.82*
	(0.14)	(0.30)	(0.07)
Log likelihood	-285.32	-184.16	-900.99
Whole Wheat Flour			
α_1, Hodgson Mill	4.56*	2.69*	2.28*
	(0.50)	(0.80)	(0.24)
α_2, King Arthur	3.61*	-0.41	2.46*
	(0.49)	(1.29)	(0.24)
α_3, Gold Medal	5.78*	2.69*	2.57*
	(0.53)	(0.78)	(0.25)
α_4, GO Organic	4.70*	1.58	2.45*
	(0.51)	(0.90)	(0.24)
α_{price}	-1.80*	-1.93*	-0.97*
	(0.16)	(0.34)	(0.07)
Log likelihood	-261.17	-140.38	-882.66
Number of Observations	235	230	215

[a] One asterisk (*) denotes values that are statistically significant at the 0.05 level.
[b] Numbers in parentheses are standard errors.

Table II-2. IAL and RPL Models Estimates by Product Category and Elicitation Method

	Dishwashing Liquids						Ground Beef						Whole Wheat Flour					
	IAL			RPL			IAL			RPL			IAL			RPL		
	Hyp	Real	Rank	Hyp	Real	Rank	Hyp	Real	Rank	Hyp	Real	Rank	Hyp	Real	Rank	Hyp	Real	Rank
Utility function																		
α_1	-12.24*[a]	1.56	2.65*	13.32*	-5.60	8.52*	0.30	1.31	1.25	29.42*	6.31	5.82*	-11.05*	8.22*	5.98*	17.11*	21.25	4.51*
	(0.72)[a]	(0.93)	(0.74)	(2.49)	(17.55)	(0.00)	(0.93)	(0.95)	(0.77)	(10.12)	(8.85)	(0.00)	(1.18)	(1.66)	(1.23)	(4.53)	(12.94)	(.0.00)
α_2	-12.13*	2.04*	2.04*	11.04*	-25.46	5.42*	3.60*	3.81*	3.52*	40.29*	14.31*	7.09*	-11.86*	1.20	1.47*	12.35*	-16.92	2.85*
	(0.74)	(0.84)	(0.87)	(2.08)	(29.02)	(0.85)	(1.14)	(0.82)	(0.80)	(13.95)	(6.62)	(0.00)	(1.12)	(1.58)	(0.67)	(4.36)	(5D+8)	(0.00)
α_3	-12.68*	2.57*	2.89*	10.38*	13.88	4.94*	0.84	3.80*	3.01*	27.05*	12.96*	4.90*	-7.80*	4.84*	2.40*	21.86*	-1.26	2.95*
	(0.74)	(0.97)	(0.86)	(1.96)	(10.37)	(0.00)	(1.00)	(0.95)	(0.76)	(9.49)	(5.68)	(0.00)	(1.51)	(1.27)	(0.66)	(5.61)	(7.11)	(0.00)
α_4	-12.93*	5.50*	8.35*	11.80*	18.09	3.35*	-0.13	6.37*	7.44*	1.75	20.30*	4.52*	-11.02*	2.62*	1.26*	17.39*	8.12	7.53*
	(0.67)	(0.81)	(2.44)	(2.44)	(14.73)	(0.00)	(0.89)	(0.80)	(0.76)	(4.76)	(8.27)	(0.01)	(1.15)	(1.28)	(0.59)	(4.48)	(11.22)	(0.00)
α_{price}	-1.64*	-1.62*	-1.52*	-4.37*	-15.84	-2.41*	-1.85*	-1.92*	-1.43*	-10.26*	-12.12*	-2.19*	-2.91*	-2.42*	-1.08*	-6.86*	-14.67	-2.04*
	(0.22)	(0.37)	(0.31)	(0.77)	(11.30)	(0.26)	(0.28)	(0.35)	(0.27)	(3.37)	(4.25)	(0.00)	(0.54)	(0.55)	(0.22)	(1.69)	(8.61)	(0.00)
Availability/standard deviation parameters[c]																		
β_1/σ_1	17.37*	22.87*	19.61*	5.07*	27.48	4.43*	16.38*	22.30*	18.76*	16.39*	11.96*	4.56*	16.42*	-2.11*	-2.53*	8.13*	13.97	4.79*
	(0.11)	(0.07)	(0.08)	(1.08)	(24.15)	(0.19)	(0.09)	(0.07)	(0.08)	(6.43)	(5.78)	(0.00)	(0.17)	(0.25)	(0.29)	(2.64)	(7.84)	(0.00)
β_2/σ_2	17.49*	22.39*	19.58*	4.55*	41.42	3.71*	16.85*	20.24*	18.69*	11.85*	19.60*	6.13*	16.43*	19.30*	18.53*	8.50*	912.43	8.08*
	(0.08)	(0.07)	(0.08)	(1.16)	(34.05)	(0.58)	(0.08)	(0.07)	(0.08)	(4.43)	(6.79)	(0.00)	(0.08)	(0.07)	(0.07)	(2.42)	(2D+8)	(0.00)
β_3/σ_3	-17.14*	-19.65*	-18.67*	4.71*	27.64	7.16*	-16.28*	-19.78*	-17.94*	14.92*	13.95*	6.12*	-16.39*	-18.55*	-17.89*	7.08*	30.17	4.74*
	(0.08)	(0.07)	(0.08)	(1.62)	(21.10)	(0.00)	(0.09)	(0.07)	(0.08)	(5.81)	(4.97)	(0.00)	(0.10)	(0.07)	(0.07)	(1.87)	(19.96)	(0.00)
β_4/σ_4	16.90*	-2.55*	-1.91*	6.21*	20.07	3.01*	16.02*	-2.25*	-1.82*	31.90*	6.50*	4.91*	16.58*	21.04*	18.53*	7.41*	15.89	2.96*
	(0.18)	(0.31)	(0.23)	(1.46)	(13.95)	(0.00)	(0.18)	(0.27)	(0.22)	(13.04)	(2.41)	(0.02)	(0.18)	(0.07)	(0.07)	(2.45)	(11.78)	(0.00)
β_{None}	-1.61*	17.28*	17.06*				-1.91*	17.63*	17.23*				-1.31*	16.85*	17.07*			
	(0.18)	(0.22)	(0.16)				(0.23)	(0.44)	(0.20)				(0.16)	(0.11)	(0.12)			
LL[d]	-167.69	-24.11	-103.45	-223.51	-106.05	-672.07	-154.09	-49.92	-126.12	-174.39	-118.56	-647.10	-155.24	-4.73	-105.24	-201.02	-85.24	-729.44
# of obs.	235	230	215	235	230	215	235	230	215	235	230	215	235	230	215	235	230	215

[a] One asterisk (*) denotes values that are statistically significant at the 0.05 level.
[b] Numbers in parentheses are standard errors.
[c] Estimates are availability parameters in the IAL standard deviations in the RPL, respectively.
[d] LL is log likelihood value.

Table II-3. Market Share Estimates from Experiments and the Actual Field by Econometric Models and Elicitation Methods

	MNL			IAL			RPL			
	Hyp	Non hyp	Rank	Hyp	Non hyp	Rank	Hyp	Non hyp	Rank	Grocery Store
Dishwashing Liquid										
Dawn	0.63	0.48	0.50	0.47	0.37	0.57	0.56	0.29	0.57	0.50
	(0.05)[a]	(0.10)	(0.03)	(0.07)	(0.11)	(0.08)	(0.09)	(0.18)	(0.00)	
Joy	0.30	0.37	0.34	0.52	0.60	0.31	0.32	0.18	0.23	0.37
	(0.04)	(0.09)	(0.02)	(0.07)	(0.10)	(0.08)	(0.09)	(0.12)	(0.00)	
Palmolive	0.06	0.13	0.13	0.00	0.00	0.00	0.07	0.35	0.19	0.14
	(0.02)	(0.05)	(0.01)	(0.00)	(0.00)	(0.00)	(0.03)	(0.15)	(0.00)	
Eco-Plus	0.02	0.02	0.03	0.01	0.03	0.12	0.05	0.14	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.02)	(0.06)	(0.00)	
Ground Beef										
Fresh	0.31	0.22	0.35	0.07	0.15	0.14	0.33	0.30	0.34	0.33
	(0.04)	(0.10)	(0.03)	(0.04)	(0.08)	(0.05)	(0.09)	(0.18)	(0.00)	
Lean	0.58	0.52	0.41	0.93	0.83	0.76	0.54	0.45	0.40	0.41
	(0.04)	(0.08)	(0.02)	(0.04)	(0.10)	(0.06)	(0.08)	(0.11)	(0.00)	
Diet Lean	0.11	0.25	0.19	0.00	0.00	0.00	0.11	0.23	0.21	0.26
	(0.02)	(0.06)	(0.02)	(0.00)	(0.00)	(0.00)	(0.02)	(0.08)	(0.00)	
Organic	0.01	0.01	0.05	0.00	0.02	0.11	0.03	0.02	0.06	0.00
	(0.00)	(0.01)	(0.01)	(0.00)	(0.11)	(0.03)	(0.02)	(0.05)	(0.00)	
Whole Wheat Flour										
Hodgson Mill	0.13	0.32	0.27	0.93	0.11	0.07	0.24	0.63	0.28	0.21
	(0.03)	(0.08)	(0.02)	(0.06)	(0.00)	(0.00)	(0.08)	(0.18)	(0.00)	
King Arthur	0.01	0.00	0.12	0.02	0.18	0.52	0.02	0.00	0.20	0.16
	(0.00)	(0.00)	(0.01)	(0.02)	(0.15)	(0.08)	(0.05)	(0.06)	(0.00)	
Gold Medal	0.83	0.63	0.50	0.00	0.00	0.00	0.69	0.28	0.22	0.28
	(0.04)	(0.08)	(0.03)	(0.00)	(0.00)	(0.00)	(0.15)	(0.08)	(0.00)	
GO Organic	0.03	0.05	0.12	0.05	0.72	0.41	0.06	0.06	0.31	0.35
	(0.01)	(0.01)	(0.01)	(0.04)	(0.15)	(0.08)	(0.06)	(0.16)	(0.00)	

[a] Numbers in parentheses are standard errors determined via parametric bootstrapping.

Table II-4. Prediction Performance by Product Category, Elicitation Methods, and Econometric Models

	MNL		IAL		RPL	
	MSE[a]	OSLLF[b]	MSE	OSLLF	MSE	OSLLF
Dishwashing Liquid						
Hypothetical	0.028	-1.065	0.043	-3.336	0.014	-1.054
	[0.006,0.078][c]	[-1.142,-1.028]	[0.027,0.133]	[-3.488,-3.238]	[0.003,0.129]	[-1.253,-1.016]
Nonhypothetical	0.001	-1.006	0.088	-3.597	0.124	-1.298
	[0.001,0.090]	[-1.141,-1.002]	[0.026,0.277]	[-3.863,-3.483]	[0.013,0.444]	[-2.615,-1.035]
Real ranking	0.002	-1.019	0.041	-3.517	0.032	-1.052
	[0.001,0.010]	[-1.035,-1.012]	[0.026,0.116]	[-3.641,-3.386]	[0.032,0.032]	[-1.052,-1.052]
Ground Beef						
Hypothetical	0.053	-1.208	0.407	-6.108	0.040	-1.196
	[0.027,0.098]	[-1.293,-1.151]	[0.267,0.482]	[-6.721,-5.580]	[0.030,0.106]	[-1.321,-1.173]
Nonhypothetical	0.027	-1.139	0.281	-6.216	0.004	-1.126
	[0.002,0.118]	[-1.343,-1.097]	[0.113,0.384]	[-6.990,-6.108]	[0.003,0.172]	[-2.351,-1.098]
Real ranking	0.007	-1.143	0.239	-5.876	0.005	-1.140
	[0.004,0.016]	[-1.164,-1.130]	[0.133,0.335]	[-6.210,-5.623]	[0.005,0.005]	[-1.140,-1.139]
Whole Wheat Flour						
Hypothetical	0.441	-2.521	0.700	-6.257	0.251	-1.873
	[0.343,0.524]	[-2.877,-2.212]	[0.425,0.809]	[-6.949,-5.709]	[0.046,0.655]	[-7.422,-1.428]
Nonhypothetical	0.251	-2.413	0.221	-6.084	0.191	-1.871
	[0.228,0.398]	[-3.330,-2.244]	[0.163,0.375]	[-6.336,-6.067]	[0.047,0.425]	[-11.22,-1.676]
Real ranking	0.105	-1.561	0.229	-6.053	0.010	-1.364
	[0.074,0.143]	[-1.641,-1.491]	[0.185,0.369]	[-6.184,-5.973]	[0.010,0.010]	[-1.364,-1.364]
Total						
Hypothetical	0.522	-4.794	1.150	-15.701	0.305	-4.124
Nonhypothetical	0.279	-4.558	0.590	-15.897	0.319	-4.295
Real ranking	**0.114**	**-3.722**	**0.509**	**-15.447**	**0.047**	**-3.556**

[a] MSE is mean squared error between predicted and actual market share summed across each product

[b] OSLLLF is the estimated likelihood function value observed at actual market share values.

[c] Numbers in brackets are 95% confidence intervals determined via parametric bootstrapping.

Table II-5. F-Statistics from ANOVA Tests for Effect of Product Category, Elicitation Method, and Econometric Model on Prediction Performance

Variable	Degrees of Freedom	Dependent Variable MSE[a]	Dependent Variable OSLLF[b]
Product Category	2	33.11*[c]	162.31*
Elicitation Method	2	13.52*	4.79*
Econometric Model	2	22.52*	1481.89*
Product Category × Elicitation Method	4	7.81*	2.6
Product Category × Econometric Model	4	6.28*	58.28*
Elicitation Method × Econometric Model	4	2.01	0.5
R^2		0.96	0.99
Number of Observations		27	27

[a] Dependent variable is mean squared error between predicted and actual market shares. Reported values are F-statistics associated with the null hypothesis that the row variable does not affect mean squared error.
[b] Dependent variable is the log likelihood function evaluated at out of true market share values. Reported values are F-statistics associated with the null hypothesis that the row variable does not affect OSLLF.
[c] One asterisk indicates the null hypothesis of no effect of the independent variable can be rejected at 0.05 level or lower.

CHAPTER III

FAIRNESS AND FOOD CHOICE

Introduction

Although standard models of individual decision making have historically relied on the assumption of self-interest, experimental economists have generated a wealth of evidence that people are motivated by altruism, equity, and reciprocity, at least in certain circumstances (e.g., see Andreoni and Miller 2002; Güth, Schmittberger, and Schwarze 1982; Fehr, Kirchsteiger, and Ridel 1993). Such findings have led to the development of models which seek to incorporate other-regarding behavior into the traditional economic framework. Such models assume that in addition to their own payoffs, people also care about the payoffs others receive. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), for examples, proposed alternative models of self-centered inequality aversion. Both models posit that people are motivated by their own payoff and their payoff relative to the payoffs of others. By contrast, Charness and Rabin (2002) proposed a model where, in addition to self-interest, people exhibit preferences for efficiency and maximizing minimum payoffs. Although such fairness models have recently gained much attention, there have only been a few attempts to compare the relative explanatory power of these theories (e.g., see Bereby-Meyer and Niederle 2005; Engelmann and Strobel 2004; Fehr, Naef, and Schmidt 2006; Bolton and Ockenfels 2006). To date, however, such comparisons have primarily been limited to abstract, experimental games devoid of naturally occurring field context. Furthermore, although there are some counter-examples, the vast majority of such experiments have been conducted with a convenient sample of student subjects. It is unclear whether and to what extent other-regarding behavior will hold up when the decision context is moved to a more natural setting or when money allocations are no longer anonymous (e.g., see the findings in List, 2006). Finally, it is of interest to determine whether the measured preference for fairness exhibit predictive validity in the sense that they explain choice in a different, but related context.

We consider the explanatory power of several models of other-regarding behavior in a field context for which all people have experience: food choice. Recent years have witnessed pronounced differentiation of food products, ranging from organic to "eco-friendly" to "hormone free" food products. The recent growth in the organic food market is often attributed, in part, to people's concerns about inequity in the food supply chain and motivations to support small farmers. Indeed, one of key principles of organic agriculture is the concern of fairness which emphasizes the relationships to between all parties in the food chain – farmers, processors, distributors, traders and consumers (International Federation of Organic Agriculture Movements). Despite the arguments by some that "fairness" and support for small farms is a key benefit of organic products, we are unaware of any empirical research actually linking fairness motivations with consumer demand for organic food.

In this paper, we seek to determine whether fairness considerations carry over to food choice and determine the extent to which the fairness models proposed in the literature explain food purchasing behavior. Moreover, we elicit consumers' perceptions about the distribution of benefits resulting from the sales of non-organic and organic food and determine the extent to which preferences for the distribution of benefits can explain preferences for organic food. Our results indicate that the fairness models proposed in the literature do not exhibit much explanatory power unless modified in nontrivial ways. Finally, we find that preferences for distribution of benefits, along with measured beliefs about the relative distribution on benefits accruing to producers of organic and conventional foods, are significant factors explaining consumer willingness to pay a premium for organic food.

Methods and Procedures

Survey Design

A mail survey was developed and sent to a random sample of 2,000 consumers in the United States in April 2007.[1] The survey contained a stated-preference experiment in which people were asked to respond to a series of purchase intention questions. In particular, respondents were asked to indicate how likely they were to purchase 12 loaves of bread that differed by price and the amount of profit from the purchase going to the following participants in the food marketing channel: small farmers (defined as farming less than 500 acres), large farmers (defined as farming 500 acres or more), agribusiness processors (such as wheat millers and bakers), and supermarkets. For each loaf of bread, people were asked to indicate how likely they were to buy the loaf on a scale of 0 to 10, where 0 was defined as "definitely would not buy," 5 was defined as "equal chance of buying and not buying," and 10 was defined as "definitely would buy." Across the 12 bread options, prices were varied among the values of $1.99, $2.99, and $3.99 and the profits accruing to each of the participants in the food marketing channel was varied among the values of $0.01, $0.07, and $0.15. Thus, the experimental design consisted of price being varied at three levels and the profits accruing to the four participants in the production of bread (small farmers, large farmers, agribusiness processors, and supermarkets) each being varied at three levels, creating $3^5 = 243$ possible types of bread. Each survey contained 12 descriptions of bread that were randomly selected from the full set of 243 (i.e., no two surveys were alike: each survey had a different set of 12 bread options). An example of two bread options presented to one respondent is shown in figure III-1.

These stated preference questions were designed to determine people's preferences for profits accruing to different participants in the food supply chain. An additional goal of this study is to determine the predictive validity of such fairness measures and determine whether such preferences also relate to people's willingness-to-pay (WTP) a premium for organic vs. non-organic bread. As such, the survey also elicited people's beliefs about the distribution of benefits across the supply channel resulting from the sale of organic and non-organic loaves of bread. People were asked to indicate how much they thought each of the following participants in the supply chain, as a whole, profited from the sale of a single loaf of organic and non-organic bread: small farmers, large farmers, agribusiness processors, and supermarkets. Participants responded by checking a box with competing dollar amounts between \$0.01 to \$0.05, \$0.06 to \$0.10, and \$0.11 to \$0.15.

To determine the extent to which the measured preferences for distributions of outcomes across the food supply chain can explain people's food preferences, the survey contained a question where people were asked to indicate the maximum premium they would be willing to pay for an organic loaf of bread over a conventional loaf of bread, assuming both were the same brand name. Finally, as will be explained in more detail in a subsequent sub-section, to determine the extent to which fairness concerns motivate people's willingness-to-pay a premium for organic bread, several additional questions were asked. In particular, respondents were asked to state: (a) how much they would normally expect to pay for a single loaf of conventional and organic bread (i.e., expected market prices) and (b) the most they would be willing to pay for a conventional and organic loaf of bread if no other market participant (farmers, agribusinesses, or grocery store) made any profit and they were the only one that benefited from the purchase.

Fairness Models and Econometric Methods

To illustrate the general modeling approach, first ignore preferences for equality and assume that individual *i*'s utility from the purchase a loaf option *j* is:

$$U_{ij} = \alpha_j - \beta P_{ij} + \delta_1 \pi_{SF,ij} + \delta_2 \pi_{LF,ij} + \delta_3 \pi_{AB,ij} + \delta_4 \pi_{GS,ij}, \tag{3.1}$$

where P_{ij} is the price of the j^{th} loaf of bread, π_{SF}, π_{LF}, π_{AB}, π_{GS} are profits to each of the members of the food supply chain (small farmers, large farmers, agribusiness processors, and grocery stores, respectively), α_j is and alternative specific constant related to the utility of having and not-having a loaf of bread, and β denotes the marginal utility of income.

Because the fairness models introduced by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) involve a comparison of benefits to self with benefits to others, it is useful to re-write (3.1) to determine the net benefit a consumer receives from the purchase of bread option *j*. In particular, we calculate the "selfish" consumer surplus of an option by finding the price level, *P*, that makes a consumer indifferent to the buying when $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$, meaning all other parties in the supply chain receive no benefit. Without loss of generality, assume people are indifferent to buying and not buying when $U_{ij} = 0$. Thus, willingness-to-pay for the loaf of bread is determined as $WTP = \alpha_j/\beta$ or $\alpha_j = \beta WTP$. Substituting this expression into equation (3.1) and rearranging yields:

(3.2) $$U_{ij} = \beta(WTP - P_{ij}) + \delta_1 \pi_{SF,ij} + \delta_2 \pi_{LF,ij} + \delta_3 \pi_{AB,ij} + \delta_4 \pi_{GS,ij},$$

where ($WTP - P_{ij}$) represents the consumer's benefit or consumer surplus from the purchase of option *j*. It is important to note that, following Train and Weeks (2005), *WTP* is a coefficient directly estimatable from the respondent's choices, and as such, the consumer benefit (or consumer surplus) from purchasing an option can be determined as the difference between estimated willingness-to-pay and price. Given this framework, we let $\pi_{C,ij}$ denote the *i*th consumer's benefit or "profit" from bread option *j*, (WTP- P_{ij}).

To determine the extent to which the fairness considerations explain food purchase behavior, we modify equation (3.2) as follows:

(3.3) $$U_{ij} = \beta\pi_{C,ij} + \delta_1 \pi_{SF,ij} + \delta_2 \pi_{LF,ij} + \delta_3 \pi_{AB,ij} + \delta_4 \pi_{GS,ij} + \lambda Fair_{ij},$$

where the variable $Fair_{ij}$ corresponds to a particular notion of fairness or inequality aversion proposed in the literature and described below.

First, we consider the model of inequality aversion advocated by Fehr and Schmidt (1999, henceforth FS). One portion of their inequality aversion motive can be captured as follows:

$$FSa_{ij} = -\frac{1}{4}[\max(\pi_{SF,ij} - \pi_{C,ij},0) + \max(\pi_{LF,ij} - \pi_{C,ij},0) + \max(\pi_{AB,ij} - \pi_{C,ij},0) + \max(\pi_{GS,ij} - \pi_{C,ij},0)] \tag{3.4}$$

If FSa_{ij} in equation (3.4) is substituted into the variable $Fair_{ij}$ in equation (3.3), then λ provides a measure of people's aversion to disadvantageous inequity. The FS model also proposes that people are averse to being in an advantageous position as well. Thus, the other inequality aversion motive can be captured as follows:

$$FSb_{ij} = -\frac{1}{4}[\max(\pi_{C,ij} - \pi_{SF,ij},0) + \max(\pi_{C,ij} - \pi_{LF,ij},0) + \max(\pi_{C,ij} - \pi_{AB,ij},0) + \max(\pi_{C,ij} - \pi_{GS,ij},0)] \tag{3.5}$$

If FSb_{ij} in equation (3.5) is into substituted the variable $Fair_{ij}$ in into equation (3.3), then λ provides a measure of people's aversion to advantageous inequity. If one includes a linear term for the payoff to each participant in the supply chain, as in equation (3.3), then the advantageous and disadvantageous motives are not separately identified, i.e., there is a linear identity between the linear payoff terms and the sum of the advantageous and disadvantageous measures in equations (3.4) and (3.5). Thus, following Engelmann and Strobel (2004), we investigate the ability of the FS model to explain behavior by substituting the following expression into equation (3.3)

$$Fair_{ij} = FSa_{ij} + FSb_{ij} . \tag{3.6}$$

Secondly, we consider the model of equity, reciprocity, and competition proposed by Bolton and Ockenfels (2000, henceforth ERC) which assumes people dislike a difference between their own payoff and the average payoff. The ERC motivation can be expressed as

(3.7) $$Fair_{ij} = ERC_{ij} = -100 \times \left| \frac{1}{5} - \frac{\pi_{C,ij}}{EFF_{ij}} \right|,$$

where $EFF_{ij} = \pi_{C,ij} + \pi_{SF,IJ} + \pi_{LF,ij} + \pi_{AB,ij} + \pi_{GS,ij}$, which represents the sum of profits to all five participants in the supply chain, including the consumer. Equation (3.7) shows that people prefer the average profit to be as close as possible to their own profit. The key contrast between the FS and ERC models is that in the ERC model, people are happy if they receive the average payoff regardless of whether some people earn more or less, but the same is not necessarily true in the FS model. In the FS model, individual *i* can be made better off by re-distributing payoffs.

The third model we consider is motivated by Charness and Rabin (2002), who posit that people are motivated, in part, by total surplus or efficiency. Thus, to implement the Charness and Rabin (2002) model, we simply substitute the following expression into equation (3.3):

(3.8) $$Fair_{ij} = EFF_{ij}$$

Finally, we explore an intuitive notion of inequality aversion, and investigate whether people are concerned about the standard deviation of profits across all supply chain participants,

(3.9) $$Fair_{ij} = SD_{ij} = -\sqrt{\frac{1}{5}\sum_{k}(\pi_{kj} - \frac{EFF_{ij}}{5})^2}$$

Because the standard deviation provides a depiction of profit variability, it can be interpreted as an alternative motivation for inequality aversion.

The question we now consider is how to empirically estimate the parameters of the model in equation (3.3). First, we adopt the random utility framework popularized by McFadden, and assume that the indirect utility function given in (3.3), while known by the respondent, is only observable to the analyst with error. Second, given that the question format requested people to respond with a rating on a cardinal scale (where the rating directly corresponds to the chance of making a purchase – i.e., 0=0% chance of making a purchase, 1=10% chance of making a purchase, etc.), we assume that an ordinary least squares model is appropriate, where the person's rating is a proxy for the utility derived from an option, i.e., $U_{ij} = Rating_{ij} + \varepsilon_{ij}$. Third, because each respondent was asked to answer 12 questions regarding how likely they were to buy a loaf of bread, the random errors are unlikely to be independent within-subject. Stated differently, unobservable heterogeneity is likely to be present across subjects. Indeed, Erlei (2008) recently pointed out the importance of heterogeneity of preferences and showed that it plays an important role in understanding laboratory behavior. Taken together, these considerations imply the following empirical model:

$$\begin{aligned} Rating_{ij} - 5 &= \beta(WTP - P_{ij}) \\ &\quad + \delta_1 \pi_{SF,ij} + \delta_2 \pi_{LF,ij} + \delta_3 \pi_{AB,ij} + \delta_4 \pi_{GS,ij} + \lambda Fair_{ij} + u_i + \varepsilon_{ij} \end{aligned} \tag{3.10}$$

where $Fair_{ij}$ is one of the measures in equations (3.6), (3.7), (3.8), or (3.9), $u_i \sim N(0, \sigma_u^2)$ is a subject-specific random effect, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is a random error term. Note that in (3.10), the constant 5 has been subtracted from the dependent variable. This step was taken so that the estimated value, *WTP*, corresponds to the dollar amount that makes people indifferent to buying a loaf of bread when all other parties receive profit equal to zero (by construction, a rating of 5 implies indifference toward buying, i.e., 50% chance of buying).

To test the relative performance of the four fairness models outlined above, i.e., equations (3.6), (3.7), (3.8), or (3.9), we consider how well the models performed in predicting out of sample by using cross-validation. In particular, the sample was randomly split in half, and each of the four fairness models were estimated using one half of the data. Then we investigated how well the estimated models predicted the "hold out" sample, which was the other half of the data not used in the estimation. This process was then repeated by switching the estimation and hold-out samples. To judge out-of-sample prediction performance, two model selection criteria were used: mean squared error (MSE) and the out-of-sample log likelihood function (OSLLF) approach. The MSE is simply the average of the squared difference between the estimated rating and the actual rating of the desirability of each bread option. A model with a lower MSE is preferred. The OSLLF method ranks models by likelihood function values observed at out of sample observations. The OSLLF selects the model with the highest out-of-sample log likelihood function value and is calculated as:

(3.11) $$OSLLF = \sum_{i=1}^{N/2}\sum_{j=1}^{J} \log \phi(Z_{ij}),$$

where $\phi(Z_{ij}) = \frac{1}{\sqrt{2\pi\varphi}} \exp\left\{ -\frac{(Rating_{ij}^{H} - \hat{Rating}_{ij}^{E})^2}{2\varphi} \right\}$ is the probability density function for an out-of-sample observation, $Rating_{ij}^{H}$ is the actual rating of option *j* in the "hold-out" or out-of-sample dataset, and $\hat{Rating}_{ij}^{E}$ is the predicted rating of option *j* using parameters obtained by fitting a model to the "estimation" data set.

Estimating the Fairness-Induced Willingness-to-Pay Premium for Organic Food

In addition to investigating consumer concerns for distribution of benefits across the food supply chain, it is also of interest to determine the extent to which such considerations explain people's preferences for organic food over conventional food. Once the coefficients for a given fairness model have been estimated, they can be combined with people's stated WTP, people's perceived prices for organic and non-organic bread, and people's beliefs about the profit levels for each participant in the supply chain to calculate the predicted utility for organic and non-organic bread:

$$(3.12) \qquad \hat{U}_i^O = \hat{\beta}(WTP_i^O - P_i^O) + \hat{\delta}_1\pi_{SF,i}^O + \hat{\delta}_2\pi_{LF,i}^O + \hat{\delta}_3\pi_{AB,i}^O + \hat{\delta}_4\pi_{GS,i}^O + \hat{\lambda}Fair_i^O$$

and

$$(3.13) \qquad \hat{U}_i^C = \hat{\beta}(WTP_i^C - P_i^C) + \hat{\delta}_1\pi_{SF,i}^C + \hat{\delta}_2\pi_{LF,i}^C + \hat{\delta}_3\pi_{AB,i}^C + \hat{\delta}_4\pi_{GS,i}^C + \hat{\lambda}Fair_i^C$$

where the *O* and *C* superscripts denote organic and conventional bread, respectively, WTP_i^k is individual *i*'s stated willingness-to-pay for the k^{th} type of bread when no other party in the food supply chain profits, P_i^k is *i*'s stated belief about what they would pay for the k^{th} type of bread in the grocery store, and where $\pi_{t,i}^k$ is the *i*'s stated belief about the profits accruing to the t^{th} business type for the k^{th} type of bread. With equations (3.12) and (3.13), the premium people are willing to pay for organic bread can be determined by finding the price difference between organic and conventional bread, (P^O – P^C), that makes a person indifferent to purchasing organic or conventional non-organic bread:

$$(3.14)\qquad \begin{aligned}(P_i^O - P_i^C) &= (WTP_i^O - WTP_i^C)\\ &\quad -\frac{1}{\hat{\beta}}\Big[\hat{\delta}_1(\pi_{SF,i}^C - \pi_{SF,i}^O) + \hat{\delta}_2(\pi_{LF,i}^C - \pi_{LF,i}^O)\\ &\quad + \hat{\delta}_3(\pi_{AB,i}^C - \pi_{AB,i}^O) + \hat{\delta}_4(\pi_{GS,i}^C - \pi_{GS,i}^O) + \hat{\lambda}(Fair_i^C - Fair_i^O)\Big]\end{aligned}$$

This price difference can be interpreted as the estimated premium consumers are predicted to be willing to pay for organic bread over conventional non-organic bread. This measure can be decomposed into two parts. The first term in the right hand side of the equality (3.14) represents the organic premium motivated by concerns unrelated to the distribution of payouts accruing to parities in the food supply chain such as concerns about the environment, health, quality, and etc., and the second term represents the organic premium explained solely by concerns for payoffs accruing to other participants in the supply chain. Thus, we can investigate how much the preferences for the distribution of benefits drive WTP for organic vs. non-organic bread. The portion of the willingness-to-pay premium for organic bread explained by payoff-induced motives is:

$$\frac{-\frac{1}{\hat{\beta}}\Big[\hat{\delta}_1(\pi_{SF,i}^C - \pi_{SF,i}^O) + \hat{\delta}_2(\pi_{LF,i}^C - \pi_{LF,i}^O) + \hat{\delta}_3(\pi_{AB,i}^C - \pi_{AB,i}^O) + \hat{\delta}_4(\pi_{GS,i}^C - \pi_{GS,i}^O) + \hat{\lambda}(Fair_i^C - Fair_i^O)\Big]}{(P_i^O - P_i^C)}$$

Finally, to investigate external validity, we simply calculate the correlation coefficient between the predicted willingness-to-pay premium for organic bread given in equation (3.14) and the person's actual stated willingness-to-pay premium for organic bread obtained in the survey.

Results

Overall 210 completed surveys were returned. After accounting for undeliverable addresses, a response rate of 11.5% is implied. Although the response rate is somewhat low, we emphasize that we were able to obtain data from a much more diverse subject pool than what would have been the case had a convenience sample of students been used. For example, 62% of respondents were female (which is likely a result of the fact that we asked the primary food shopper in the household to complete the survey), the mean age of the respondents was about 56 years old (with a standard deviation of 15 years), 55% of the sample had earned a bachelor's degree, 17% had children under the age of 12 in the household, and only a small fraction of the sample (15%) said they or someone in their immediate family farmed or ranched for a living. Importantly, we do not claim that our estimates of preferences for relative payoffs are representative of the U.S. population *per se* but rather ask, for this sample of people, whether they exhibit preferences for the distribution of payoffs across the food supply chain and test whether these measured preferences related the willingness-to-pay a premium for organic food.

Table III-1 reports estimates for each of four fairness models. For each model, except for the FS model, the coefficient β is positive, meaning the marginal utility of income is positive and consumers care about their own benefit or "profit." All model specifications indicate people are willing to pay about \$1.45 for a loaf of bread assuming no other participant in the supply chain benefits from the sale. Coefficients for payoffs to small farmers, α_1, are positive and statistically significant in each model, meaning people primarily care about the benefits to small farmers.

The key results relate to the estimate of the parameter λ, which correspond to the various fairness concerns. The only fairness motivation that was statistically significant was in the ERC model, but here the coefficient was of the opposite sign than expected – i.e., people preferred receiving payouts the diverged from the average payout. These findings indicate that, after holding constant the payoffs to each participant in the supply chain, choices are either unaffected by concerns for inequity or efficiency or are affected in non-intuitive ways.

Nevertheless, because the models suggest people care about the payouts to small farmers, beliefs about payoff differences might explain preferences for organic bread if people believe small farmers benefit from selling organic bread. The bottom portion of table III-1 shows the premium for organic bread over conventional bread which results from payouts accruing to different parties in the food supply chain. The portion of premium explained by differences in relative payouts is ranges from 39.7% to 48.8%. One might question why these values are so large when none of the fairness parameters are statistically significant. The answer is because people care about small farmers (and the coefficients for this particular participant are large), and because people perceive small farms to derive a large benefit from organic foods. To illustrate this, table III-2 shows people's beliefs about the payoffs to different participants in the supply chain for organic and non-organic bread resulting from the sale of a single loaf of bread. As can be seen in table III-2, people believe all participants in the food supply chain benefit more from selling organic bread, but small farmers are believed to benefit more than others. The bottom portion of table III-1 also shows that for the SD and EFF models, that the predicted willingness-to-pay premium for organic bread is positively and significantly correlated with people's actual stated willingness-to-pay.

In general, table III-1 suggests that the fairness models proposed in the literature have very little explanatory power in the food choice context, at least when one controls for the payoff to each party. One possible reason is that unlike simple distributional games, a person's own payoff is much less transparent when making a purchase as it is a result of consumer surplus – the difference between WTP and price. This lack of transparency in determining one's own benefit may cause people to be less sensitive to comparisons of self vs. others than is assumed in the models proposed in the literature. As such, we consider whether modifications to the models might improve fit. In particular, we revised the fairness motivations expressed in equations (3.5), (3.6), (3.7), and (3.8), by excluding profits to self, $\pi_{C,ij,}$ and assume that people in the FS and ERC models are concerned about inequity as it relates to small farmers, the group that table III-1 indicates respondents were most concerned about. The results of these modified models are reported in table III-3.

According to the MSE and OSLLF criteria, all models in table III-3 exhibit better out-of-sample predictive performance than their respective counterparts in table III-1. Furthermore, the estimate parameters related to fairness concerns, λ, are now all statistically significant and of expected sign. For SD model, the coefficient on the standard deviation of profits across supply chain excluding consumers' profits is 5.655, meaning consumer prefer an equal distribution of profits among the agents in marketing channel. That λ in the ERC is positive indicates consumers dislike small farms receiving payoffs that differ from the average payoff to the four other business types, and a positive λ in FS model indicates that participants dislike the payoffs to small farms diverging from payoffs to large farms, agribusinesses, and grocery stores. Finally, the positive coefficient for λ in the EFF model variable implies that, all else equal, people prefer higher total profits to the four businesses. In each of the four models shown in table III-3, 38.8% to 42.3% of total willingness-to-pay premium for organic foods can be explained by consumers' concerns for the distribution of profits. Importantly, the correlation between the estimated premium and people's stated willingness to pay premium is positive and statistically significant for all four models. Comparing the relative performance of the models indicates that, overall, the modified FS model provides the best fit to the data.

Conclusions

This study investigated the extent to which several models of other-regarding can explain people's food choices. We also sought to determine whether estimated concerns for others' payoffs can explain people's willingness to pay for organic food. Using data from a mail survey administered to 207 U.S. households, we found that although people do indeed exhibit other-regarding preferences, the existing models proposed in the literature do not exhibit much explanatory power. However, when the models are modified to account for the fact that a person's own surplus is less transparent in a food purchasing context than in simple distributional games, all models exhibit better explanatory power. The modified Fehr and Schmidt (1999) model provided the best out-of-sample prediction performance.

Finally, our results indicate that people's other-regarding preferences explain a non-trivial portion of people's willingness-to-pay a premium for organic food, and that the estimated models of other-regarding behavior exhibit reasonably high external validity as they are significantly related to people's stated willingness-to-pay a premium for organic food. Our findings indicate that concerns for inequity can be observed in a field context, though not in the same manner as in simple laboratory distributional experiments.

Notes

1. The mailing list was purchased from a reputable survey research company which randomly selected names from the white pages of the telephone directory. In designing the survey, we followed the guidance and suggestions in Dillman (2000). The survey questionnaire was mailed out with a personalized cover letter, and the mailing included a prepaid return envelope. One week after the survey was send, a reminder postcard was mailed out.

Product	Definitely Would Not Buy					Equal Chance of Buying and Not Buying					Definitely Would Buy
Price of bread loaf: **$2.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$1.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10

Figure III-1. Example Survey Questions

Table III-1. Model Estimates by Fairness Models with Self-Interest

	Models			
Parameters	SD	ERC	FS	EFF
β	1.596	1.280**	-2.404	2.261**
	(1.888)[a]	(0.055)	(3.343)	(0.841)
WTP	1.475**	1.422**	1.474**	1.474**
	(0.145)	(0.019)	(0.145)	(0.145)
α_1	14.298**	13.981**	15.301**	15.301**
	(0.956)	(0.824)	(1.180)	(1.180)
α_2	-0.993	-1.344	-	-
	(0.936)	(0.835)		
α_3	-0.554	-0.465	0.538	0.538
	(1.226)	(0.491)	(0.959)	(0.959)
α_4	-1.499	-1.768**	-0.491	-0.491
	(0.975)	(0.817)	(1.170)	(1.170)
λ^b	-0.604	-0.001*	3.732	-0.933
	(4.253)	(0.000)	(3.345)	(0.836)
σ_u^2	2.729**	2.734**	2.729**	2.729**
	(0.312)	(0.312)	(0.312)	(0.312)
Portion[c]	0.410	0.488	0.397	0.432
Correlation[d]	0.324**	-0.026	0.104	0.251**
	(0.000)	(0.727)	(0.160)	(0.001)
MSE[e]	9.609	9.596	9.610	9.609
OSLLF[f]	-5340.913	-5339.478	-5341.041	-5341.046
No. of Obs.	2,484	2,484	2,484	2,484
No. of Respondents	207	207	207	207

Note: * and ** represents statistical significance at the 10% and 5% levels, respectively.

[a] Numbers in parentheses are asymptotic standard errors.

[b] SD = -standard deviation(self, small farmer, large farmer, agribusiness, grocery store),
EFF = self + small farmer + large farmer + agribusiness + grocery store,
ERC = -100×|(1/5)-(self/EFF)|,
FS = FSa + FSb = -1/4[max(small farmer–self, 0) + max(large farmer–self, 0) + max(agribusiness–self, 0) + max(grocery store–self, 0)] -1/4[max(self-small farmer, 0) + max(self-large farmer, 0) + max(self-agribusiness, 0) + max(self-grocery store, 0)].

[c] Numbers are the trimmed mean of portion of estimated people's premiums on organic over conventional that result solely from fairness concerns versus other factors, such as safety, health, or environmental concerns by discarding the five lowest and highest values.

[d] Correlation between calculated people's premium for organic versus conventional and stated people's willingness-to-pay for organic.

[e] MSE is mean squared error between predicted and stated rate.

[f] OSLLF is the estimated likelihood function value observed at stated rate values.

Table III-2. Beliefs about the Distribution of Profits across the Food Supply Chain Resulting from the Sale of a Single Organic and Non-Organic Loaf of Bread

Supply Chain Participants	Conventional Non-Organic	Organic	Difference in Organic and Non-Organic	Percent Increase from Non-Organic to Organic
Small farmers	\$0.059	\$0.073	\$0.014	23.73%
Large farmers	\$0.079	\$0.089	\$0.010	12.66%
Agribusiness	\$0.089	\$0.094	\$0.005	5.62%
Grocery store	\$0.100	\$0.108	\$0.008	8.00%

Table III-3. Model Estimates by Fairness Model without Self-Interest

	Models			
Parameters	SD	ERC	FS	EFF
β	1.340**	1.326**	1.333**	1.328**
	(0.058)[a]	(0.057)	(0.057)	(0.058)
WTP	1.587**	1.756**	1.644**	1.474**
	(0.146)	(0.162)	(0.147)	(0.145)
α_1	14.565**	14.280**	14.686**	13.815**
	(0.819)	(0.817)	(0.820)	(0.709)
α_2	-0.854	-1.717**	-0.896	-1.486**
	(0.835)	(0.864)	(0.833)	(0.713)
α_3	1.350*	-0.620	0.789	-0.948*
	(0.736)	(0.500)	(0.574)	(0.492)
α_4	-1.113	-2.040**	-1.136	-1.977**
	(0.827)	(0.840)	(0.823)	(0.706)
λ[b]	5.655**	0.016**	5.727**	0.553*
	(1.764)	(0.005)	(1.406)	(0.301)
σ_u^2	2.741**	2.731**	2.738**	2.729**
	(0.313)	(0.312)	(0.312)	(0.312)
Portion[c]	0.405	0.423	0.415	0.388
Correlation[d]	0.300**	0.306**	0.286**	0.315**
	(0.000)	(0.000)	(0.000)	(0.000)
MSE[e]	9.557	9.592	9.537	9.609
OSLLF[f]	-5334.278	-5338.985	-5331.771	-5341.041
No. of Obs.	2,484	2,484	2,484	2,484
No. of Respondents	207	207	207	207

Note: * and ** represents statistical significance at the 10% and 5% levels, respectively.

[a] Numbers in parentheses are asymptotic standard errors.

[b] SD = -standard deviation(small farmer, large farmer, agribusiness, grocery store),
EFF = small farmer + large farmer + agribusiness + grocery store,
ERC = -100×|(1/4)-(small farmer/EFF)|,
FS = FSa + FSb = -1/3[max(large farmer–small farmer, 0) + max(agribusiness–small farmer, 0) + max(grocery store–small farmer, 0)] -1/3[max(small farmer-large farmer, 0) + max(small farmer-agribusiness, 0) + max(small farmer-grocery store, 0)].

[c] Numbers are the trimmed mean of portion of estimated people's premiums on organic over conventional that result solely from fairness concerns versus other factors, such as safety, health, or environmental concerns by discarding the five lowest and highest values.

[d] Correlation between calculated people's premium for organic versus conventional and stated people's willingness-to-pay for organic.

[e] MSE is mean squared error between predicted and stated rate.

[f] OSLLF is the estimated likelihood function value observed at stated rate values.

CHAPTER IV

SENSITIVITY OF MIXED LOGIT ESTIMATES TO MODEL SPECIFICATION, NUMBER OF REPLICATIONS, AND SOFTWARE PACKAGE: A MONTE CARLO ANALYSIS

Introduction

There have been extensive developments in discrete choice econometrics in recent years, and such models have been utilized in fields as diverse as accounting, finance, environmental and natural resource economics, marketing, and transportation (e.g., see Boxall, Englin, and Adamowicz 2003; Brownstone, Bunch, and Train 2000; Jones and Hensher 2007; Narayanan, Desiraju, and Chintagunta 2004). Since the work of McFadden (1974), the standard in multinomial discrete choice modeling has been the multinomial (or conditional) logit (MNL) model. The popularity of the MNL is due to its ease of estimation, interpretation, and calculation of probabilities. These advantageous properties, however, come at a cost as they stem from the independence of irrelevant alternatives (IIA) assumption, which is, itself, a result of the assumption that the error terms in the random utility function are independent and identically distributed across people and choice alternatives.

Although the MNL has been the standard in discrete choice econometrics for over 30 years, people have long questioned the restrictiveness of the IIA assumption and have developed a variety of competing models that generalize the MNL. One such model is the random parameter (or mixed) logit (RPL), which relaxes the IIA assumption by modeling preference heterogeneity. McFadden and Train (2000) have shown that the RPL can approximate any underlying random utility model. This model first appeared in the economics literature a little over a decade ago, and has subsequently surged in popularity. Indeed, the RPL has risen to such prominence that it has, to a large extent, become the norm in modeling choice data. This transition has happened primarily as a result of the conceptual advantages of the RPL over the MNL (i.e., it relaxes the IIA assumption), the fact that results from the RPL have straightforward interpretations as compared to some other models that relax IIA, and because the model is supported by variety of econometric software packages.

Despite these conceptual advantages, most practitioners who have estimated a RPL are aware that the estimates can be sensitive to specification and that achieving convergence is not always easy, in part because the likelihood function is not necessarily globally concave (e.g., see the discussion in Bajari, Fox, and Ryan 2007). The RPL model lacks a closed form solution, is non-linear in parameters, and requires evaluating a likelihood function with multiple integrals either by simulation or by hierarchical Bayesian methods and Gibbs sampling. Moreover, in a simulation environment, Chiou and Walker (2007) have shown the RPL model can converge with what appear to be reasonable parameter estimates even when the model is unidentified either theoretically or empirically. None of these factors suggest that the RPL should be abandoned, but such considerations suggest the need for further research on the reliability of RPL estimates

In this paper, we carry out a Monte Carlo analysis to determine the sensitivity of RPL estimates. First, we ask how accurate are RPL estimates relative to true parameter values when there is no, low, and high preference heterogeneity (i.e., no, low, and high violations of IIA), which should give practitioners some guidance on how reliable are the estimates obtained from any one particular data set. Second, we determine how the accuracy of RPL estimates varies with (a) sample size, (b) number of replications used in the simulated maximum likelihood function, and (c) econometric software package.

The last issue is becoming increasingly important. The increased demand for estimating RPL models among researchers has resulted in the further development of several econometric software packages. In practice, many researchers rely on convenient, “canned” routines provided in popular econometric packages to estimate the RPL. Given the aforementioned concerns with the RPL, it is prudent to take a step back and determine the sensitivity of RPL results to a variety of factors including software package.[1] McCullough and Vinod (1999) argue that too little attention is paid to the numerical accuracy of econometric packages and they argue that the reliability of software packages cannot be taken for granted. This is especially true in the case of the RPL where a myriad of choices must be made to implement the simulation and optimization algorithms, some explicitly acknowledged by the software package and some that are not. That is, setting the same defaults on the number of simulation replications, search algorithm, and convergence criteria would not guarantee equivalent RPL results across software packages, in part because of the lack of a closed form solution to the RPL. Clearly practitioners would benefit from more information on the relative performance of competing software packages, and in this research we compare estimates across three of the most popular packages: SAS, LIMDEP-NLOGIT, and STATA.

Our Monte Carlo results suggest a tendency for RPL estimates to exhibit bias in small samples – an effect that diminishes as sample size increases. Similarly, the variability of RPL estimates is quite high relative to the mean in small samples. The number of replications used in the simulated maximum likelihood estimation had a relatively small effect on precision. Finally, our simulation results indicate differences in performance across software package in small samples. Across all scenarios considered, the results generated by LIDMEP-NLOGIT are most accurate.

The Random Parameter Logit Model

Random utility theory posits that the level of utility individual *i* receives from choice alternative *j* is

(4.1) $$U_{ij} = \mathbf{x}_{ij}^{'}\boldsymbol{\beta}_i + \varepsilon_{ij},$$

where $\mathbf{x}_{ij}$ is a vector of observed explanatory variables describing the characteristics or attributes of alternative *j*, $\boldsymbol{\beta}_i$ is an individual-specific parameter vector, and ε_{ij} is an iid stochastic component. The RPL capture unobservable heterogeneity by modeling the distribution of $\boldsymbol{\beta}_i$ in a sample of individuals, which is characterized by a continuous probability distribution functions. Because the utility parameters, $\boldsymbol{\beta}_i$, are unobservable, the likelihood of individual i choosing alternative j is equivalent to the joint probability that

$$\text{Prob}_{ij}(j|\mathbf{\Omega}) = \int \frac{\exp(\mathbf{x}_{ij}^{'}\boldsymbol{\beta}_i)}{\sum_{k=1}^{J}\exp(\mathbf{x}_{ik}^{'}\boldsymbol{\beta}_i)} f(\boldsymbol{\beta}_i|\mathbf{\Omega})d\boldsymbol{\beta}_i, \tag{4.2}$$

where $f(\boldsymbol{\beta}_i|\mathbf{\Omega})$ is the density of $\boldsymbol{\beta}_i$ and $\mathbf{\Omega}$ is a vector of parameters characterizing the distribution of $\boldsymbol{\beta}_i$ (Train 2003). Because equation (4.2) involves a multi-dimensional integral, it lacks a closed form solution. To circumvent this problem, Train (2003) suggests that the parameters of the model, $\mathbf{\Omega}$, be estimated by simulated maximum likelihood estimation techniques. Using such methods, the probability in (4.2) is calculated by taking the average of repeated draws from $f(\boldsymbol{\beta}_i | \mathbf{\Omega})$.

The traditional method for generating such draws relies on pseudo-random number sequences. Researchers have found that the speed of convergence can be improved by using a much smaller number of deterministic Halton draws, which evenly spreads values over the unit interval by taking constant factions based on a prime number (see Train (2003) for more detailed discussion). Bhat (2001) and Train (1999, 2003) have shown that the same level of precision can be obtained with 90% fewer Halton draws as compared to traditional random draws. Although the superiority of Halton draws is widely accepted, an open question is the *number* of draws needed to adequately simulate the probabilities in the likelihood function. Clearly, computational time increases in the number of draws, and as such, researchers would prefer a smaller number of draws. However, precision may also be increasing in the number of draws. Thus, researchers would benefit from more information on the trade-off between computational time and precision.

Monte Carlo Experiments

A Monte Carlo experiment was carried out to determine how the extent of preference heterogeneity, sample size, number of simulation replications, and software package influence RPL estimates. In the experiments, a utility function is assumed and the estimated utility parameters are compared to the true values to identify how such factors influence the precision of the estimates.

It was assumed that each simulated respondent made a choice among three alternatives (J=3), each described by two attributes, x_1 and x_2. The utility derived from option j is

(4.3) $$U_{ij} = (\beta_1 + \sigma_1 v_{1i})x_{1ij} + (\beta_2 + \sigma_2 v_{2i})x_{2ij} + \varepsilon_{ij}$$

where β_n and σ_n, n=1, 2, are the true population means and standard deviations, v_{nj} are independent standard normal variables, and ε_{ij} are type I extreme value random errors. The true parameter values assumed for β_n and σ_n are shown in Tables IV-1, IV-2, and IV-3, along with estimation results. The true mean values for β_1 and β_2 were held constant across all scenarios at 2 and 5. The true values for standard deviation, (σ_1, σ_2), were varied among the values of (0, 0), (1, 1), and (2, 5), which we refer to as no, low, and high preference heterogeneity. Note that the standard deviations (0, 0) imply no preference heterogeneity and datasets generated under this assumption are the same as those assumed in the MNL.

Hypothetical datasets were constructed of size N=200, 500, and 1,000, so as to examine the impact of the sample size (i.e., number of simulated individuals) on precision of RPL estimates. In each simulation scenario, $N \times J$ values were randomly generated for x_{1ij}, x_{2ij}, and ε_{ij}, and N values were generated for v_{1j} and v_{2j}. Values for x_{nij} and v_{nj} were each drawn from independent standard normal distributions and the stochastic error of the utility function for each alternative, ε_{ij}, was generated from a type I extreme value distribution as is assumed by the RPL model. Once a data set was generated and the parameters estimated, the process was repeated 500 times using a new set of random draws.[2] Thus, for each scenario, 500 estimates of the parameters β_1, β_2, σ_1 and σ_2, were generated. The exact code used to carry out the Monte Carlo experiments in each software package is provided in appendix IV-B. In the simulation, we also considered the number of Halton draws used in the simulated maximum likelihood function. We varied the number of draws between R=100 and R=500.

The Monte Carlo experiment was conducted with three different econometric software package widely used in economic research: (a) SAS 9.1, SAS Institute Inc., (b) NLOGIT 4.0, Econometric Software, Inc., and (c) STATA 9.2, STATACorp LP. Each of these packages contains user-friendly "canned" routines to estimate the RPL models, and we use these "canned" routines in our analysis. In particular, we used the PROC MDC with MXL specification option in SAS, RPLOGIT in NLOGIT, and MIXLOGIT in STATA. The simulated data were manipulated to feed into these routines as prescribed by each package.[3] Each software package employs simulated maximum likelihood methods to estimate the RPL. In our initial analysis, we used the default search algorithm and convergence criteria set by each software package under the premise that most practitioners would use the default, but as will be discussed momentarily, we also explored the effect of such defaults on parameter estimates.

In summary, the Monte Carlo experiment varied four treatment variables: degree of heterogeneity ((0, 0), (1, 1), and (2, 5)), sample size (200, 500, and 1,000), the number of Halton draws (100 and 500), and software package (SAS, NLOGIT, and STATA), producing 3 × 3 × 2 × 3 simulation scenarios. For each simulation scenario, 500 Monte Carlo iterations were conducted as described above.[4] Because of potential problems with non-convergence of outcomes, for each of the 500 models that were estimated in a simulation scenario, we determined whether the respective software package indicated convergence had been reached. If the software package indicated a lack of convergence, we omitted the results from the comparisons. NLOGIT and SAS never indicated non-convergence across the 500 iterations for any simulation scenario, but STATA indicated non-convergence anywhere from one to twenty three times, and in such an instance, the parameters were removed from the comparison.

Results

Table IV-1 shows the outcome when there is no preference heterogeneity. Thus, one interpretation of the results in table IV-1 is that it illustrates the consequence of using RPL when in fact MNL is the true model. Results reveal that in large samples (i.e., N=1,000), the RPL converges to the true values and generates estimates of standard deviations that are close to their true values of zero. However, if we calculate the size of the test by investigating, across the 500 iterations, how often would we incorrectly conclude that there is significant taste heterogeneity, we find that such an outcome occurs 5.8%, 12.8%, and 16.8% of the time when N=200, 500, and 1,000 respectively at 5% or lower significance level in the software package NLOGIT (see table IV-A.1 in appendix

A). Thus, on the one hand, we find that there is relatively little danger in utilizing a RPL in large samples even if preference homogeneity exists. On the other hand, if it is known that preference homogeneity is present, there is a high probability of type I error. Moreover, the MNL provides a *much* more accurate depiction of preference parameters than the RPL. To illustrate this fact, the first column of results in table IV-1 shows the results from the conventional MNL when N=200 as estimated by SAS. The mean MNL estimates, 2.094 and 5.220, are closer to the true values of 2 and 5 as compared to the RPL. Furthermore, the standard deviations of MNL parameters are quite low compared to those from RPL with R=100 (e.g., 0.356 vs. 3.268 for β_1 in SAS).

A key result shown in table IV-1, which is re-enforced by the findings in tables IV-2 and IV-3 is that when sample size is small (N=200), there is a great deal of variability in the RPL parameter estimates across the 500 iterations, and results suggest a tendency for bias. For example, results in table 2 indicate that despite the fact that the true value was 5, the mean estimate for β_2 was 6.070, 5.649, and 5.567 in SAS, NLOGIT, and STATA, respectively, given N=200 and R=100. Furthermore, the standard deviations of these estimates across the 500 iterations are quite high: 8.127, 2.504, and 1.607 for SAS, NLOGIT, and STATA, respectively. Indeed, in the case of SAS when N=200 and R=100, there is a more than 30% chance of obtaining a parameter value that is more than twice the true value! Despite the magnitude of the standard deviations, we still reject the hypothesis that the mean values equal the true values when N=200 for most simulation scenarios considered.

These results suggest that when the sample size is small (say, for example, with data from economic experiments), there may be insufficient variation to model the kind of distributional information being assumed by the RPL. That is, people may be asking "too much" of their data when trying to fit a RPL to data sets of small size. Nevertheless, results suggest that increasing the sample size readily solves the problem. One exception to this statement is in regard to the case when there is low preference heterogeneity (table IV-2). The average values for the standard deviations, σ_1 and σ_2, do not converge to their true values, (1, 1) as sample size increases. In particular, σ_2 moves farther away from the true values as sample sizes increase in all three packages; however, the standard deviations of the estimates fall and the medians approach the true values.

One somewhat surprising result is that an increase in the number of Halton draws from R=100 to R=500 used in simulated maximum likelihood function does meaningfully impact estimates. In fact, one can easily find example in tables 1, 2, and 3 where smaller numbers of Halton draws produced more precise estimates. For example, under high preference heterogeneity (table IV-3), when N=1,000, the differences between the mean estimates and true values is smaller with 100 Halton draws than with 500 draws for all software packages. Of course, in repeated samples, a larger number of Halton draws must, by definition, give a more precise approximation to the integral. Thus, one way of interpreting these findings is that the effect of moving from 100 to 500 Halton draws is too small to be detected given the number of Monte Carlo interactions employed in this study.

Across all three tables, there are non-trivial differences between the three software packages. To summarize these differences, table IV-4 shows the root mean squared error (RMSE) by sample size, software package, and the number of replications in the simulation. In large sample size, all three packages yield similar performance regardless of the degree of heterogeneity and the number of simulations. However, the mean RMSE when N=200 from SAS is 55.098 and 30.002 for R=100 and R=500, respectively, which is substantially higher than the values from NLOGIT and STATA. Overall, the lowest RMSE is generated by NLOGIT.

To further summarize the results, an analysis of variance analysis (ANOVA) was conducted to test whether RMSE is influenced by sample size, number of replications, and software package. Results in table IV-5 indicate that sample size, software package, and their interactions significantly affect the mean of RMSEs. The highest F-value corresponds to sample size; meaning we can be most confident increasing sample size improves accuracy. Table IV-5 also reveals that accuracy of RPL estimates depends on choice of software package. Number of Halton replications did not significantly affect RMSE.

What could be the causes of the differences between econometric packages? In the results in tables IV-1, IV-2, and IV-3, we utilized the default optimization algorithm and convergence criteria for each practitioner under the assumption that most practitioners would do the same. Although all three software packages utilize simulated maximum likelihood to estimate the RPL model, there are some differences in defaults employed across software packages as indicated in table IV-6. Thus, one possible explanation for the differences in performance across packages might be the different search algorithms used as defaults. For example, the default optimization techniques in SAS, NLOGIT, and STATA are quasi-Newton, BFGS, and Newton-Raphson methods, respectively. To investigate this issue, we repeated the Monte Carlo simulations but altered the search algorithms in each package. Overall, the results suggest that differences in default search algorithm, while having some effect on precision, cannot fully explain differences across software packages. Table IV-A.2 in appendix IV-A shows, for N=200 and R=100, results from altering the search technique. In SAS, the Newton-Raphson method provided slightly better estimates than those from the default, quasi-Newton, but the variability in the parameter estimates remained high relative to NLOGIT. Changing the search algorighim in NLOGIT had virtually no effect. In STATA, the default Newton-Raphson method performed better than the BFGS method, primarily as a result of one extreme outlier.

Another source of difference between packages could be the routines for generating random numbers for the Monte Carlo experiments. For example, Baiocchi (2005) compared the four random number generators in GAUSS (rndi and rndKMi functions), LIMDEP, and R through the Marsaglia's DIEHARD tests and showed that LIMDEP was superior to GAUSS in generating random numbers. To rule out the possibility that differences in the data generating process were the cause of the differences across software packages, we used generated data from one package, but estimated the model in another. For example, data was generated from SAS, but the RPL was estimated in NLOGIT. The results when there is low heterogeneity with N=200 and R=100 are shown in table IV-A.3 in appendix IV-A. Results indicate significant differences in software package even when the exact same data set is used, and in the same direction as shown in tables 1, 2, and 3, suggesting that differences in the implicit approaches used to generate random numbers are not responsible for the differences in RPL estimates across software packages.

The computation times achieved by three packages are compared in table IV-7. For all experiments, the run times by SAS are shortest; perhaps because in SAS all the data was generated at once whereas in NLOGIT and STATA, a do-loop was used to generate the data. The run times by NLOGIT were the longest for all experiments. For example, at 500 draws with *N*=1000 sample sizes, NLOGIT was eight times longer than SAS and three times longer than STATA. Thus, we find, as did McCullough and Vinod (1999) that there is a trade-off between accuracy and speed in performance by NLOGIT.

Conclusions

Although the RPL model is currently the most promising advance in modeling multinomial choices, the stability of estimates is a concerned among practitioners. Using Monte Carlo experiments, this study investigated the sensitivity of RPL estimates to changes in sample size (*N*=200, 500, and 1,000), number of replications in the simulated likelihood function (*R*=100 and 500), and econometric software package (SAS, LIMDEP-NLOGIT, and STATA) with three level of taste heterogeneity (no, low, and high). Results show that the MNL provides much more accurate estimates than the RPL when there is preference homogeneity. Moreover, in large samples there is a high likelihood of type I error – finding significant taste heterogeneity when none exists. Results indicate that when the sample size is small (i.e., *N*=200), considerable variability exists in estimated RPL parameters across 500 Monte Carlo iterations for all levels of preference heterogeneity and for each software package. However, precision increases dramatically as sample size increases. We found no relationship between the number of Halton used an accuracy of RPL estimates. Overall, NLOGIT performed relatively well for most scenarios, but came at the expense of increased computational time.

Notes

1. Although not a comparison between software packages, a few papers have compared RPL estimates across simulated maximum likelihood and Bayesian estimation methods. For example, Huber and Train (2001) compared coefficients between the simulated maximum likelihood and Bayesian methods and found that the results of two procedures are similar. In addition, Train (2001) examined the differences between the simulated maximum likelihood and Bayesian approaches in terms of run-time. All the software packages we compare in this paper rely on simulated maximum likelihood estimation.

2. A larger number of replications is often desirable in Monte Carlo studies; however, even with as few as 500 iterations, run-time was quite long with some software packages. In the most extreme case, it took 2.6 days to complete one of simulation scenario.

3. The simulated data are originally arranged with one row per choice rather than one row per alternative. In STATA and SAS, we had to manipulate the original data set such that there was one row for each alternative. In these two packages, the data sets consisted of 600, 1,500, and 3,000 rows when sample sizes are 200, 500, and 1000, respectively. NLOGIT was able to use the originally arranged data set directly.

4. The hardware used in this study was a Pentium IV (R) CPU Processors 3.40 GHz with 2 GB of RAM running on Microsoft Window XP Professional Version 2002 (Service Pack 3).

Table IV-1. Results of 500 Monte Carlo Experiments: Preference Homogeneity

		SAS-MNL	SAS-RPL			NLOGIT-RPL			STATA-RPL		
Parameter	True Value	N=200	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
R=100 Halton Draws											
β_1	2	2.094[a]	2.415	2.147	2.091	2.253	2.118	2.076	2.299	2.157	2.093
		(0.356)[b]	(3.268)	(0.277)	(0.183)	(0.488)	(0.267)	(0.183)	(0.952)	(0.283)	(0.180)
		[2.072][c]	[2.175]	[2.114]	[2.073]	[2.149]	[2.076]	[2.071]	[2.172]	[2.138]	[2.082]
σ_1	0		0.302	0.202	0.172	0.248	0.188	0.161	0.317	0.210	0.169
			(1.386)	(0.270)	(0.218)	(0.400)	(0.258)	(0.211)	(0.612)	(0.269)	(0.204)
			[0.034]	[0.040]	[0.037]	[0.025]	[0.040]	[0.033]	[0.056]	[0.063]	[0.064]
β_2	5	5.220	6.159	5.391	5.250	5.694	5.325	5.209	5.851	5.430	5.260
		(0.775)	(9.935)	(0.680)	(0.427)	(1.135)	(0.606)	(0.423)	(2.634)	(0.712)	(0.421)
		[5.144]	[5.439]	[5.324]	[5.209]	[5.495]	[5.228]	[5.187]	[5.455]	[5.338]	[5.218]
σ_2	0		0.615	0.376	0.305	0.484	0.371	0.325	0.508	0.400	0.325
			(3.702)	(0.510)	(0.397)	(0.701)	(0.494)	(0.391)	(1.177)	(0.505)	(0.376)
			[0.078]	[0.087]	[0.071]	[0.073]	[0.088]	[0.095]	[0.072]	[0.137]	[0.109]
R=500 Halton Draws											
β_1	2		2.257	2.137	2.085	2.259	2.114	2.074	2.255	2.151	2.091
			(0.483)	(0.277)	(0.183)	(0.615)	(0.267)	(0.182)	(0.504)	(0.277)	(0.180)
			[2.165]	[2.103]	[2.069]	[2.144]	[2.068]	[2.067]	[2.168]	[2.131]	[2.076]
σ_1	0		0.222	0.180	0.160	0.240	0.171	0.150	0.279	0.191	0.155
			(0.384)	(0.274)	(0.224)	(0.516)	(0.264)	(0.215)	(0.450)	(0.272)	(0.210)
			[0.007]	[0.010]	[0.010]	[0.005]	[0.006]	[0.008]	[0.012]	[0.018]	[0.016]
β_2	5		5.676	5.358	5.230	5.703	5.312	5.200	5.725	5.414	5.251
			(1.155)	(0.678)	(0.425)	(1.483)	(0.604)	(0.417)	(1.247)	(0.692)	(0.418)
			[5.402]	[5.286]	[5.185]	[5.436]	[5.223]	[5.176]	[5.468]	[5.323]	[5.213]
σ_2	0		0.369	0.302	0.242	0.443	0.321	0.284	0.418	0.351	0.290
			(0.681)	(0.506)	(0.391)	(0.822)	(0.497)	(0.399)	(0.707)	(0.510)	(0.388)
			[0.014]	[0.015]	[0.016]	[0.017]	[0.019]	[0.022]	[0.021]	[0.031]	[0.033]

[a] Mean parameter value across 500 Monte Carlo simulations.

[b] Numbers in parentheses are standard deviations of parameters across 500 Monte Carlo simulations.

[c] Numbers in brackets are median parameters across 500 Monte Carlo simulations.

Table IV-2. Results of 500 Monte Carlo Experiments: Low Preference Heterogeneity

		SAS			NLOGIT			STATA		
Parameter	True Value	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
R=100 Halton Draws										
β_1	2	2.386[a]	2.066	2.036	2.231	2.070	2.019	2.197	2.105	2.038
		(3.034)[b]	(0.310)	(0.224)	(0.990)	(0.342)	(0.228)	(0.652)	(0.365)	(0.239)
		[2.073][c]	[2.021]	[2.021]	[2.000]	[2.045]	[1.999]	[2.054]	[2.065]	[2.012]
σ_1	1	1.113	1.017	1.015	1.050	1.014	0.976	1.021	1.026	1.018
		(1.814)	(0.371)	(0.273)	(0.889)	(0.409)	(0.250)	(0.689)	(0.411)	(0.273)
		[0.939]	[1.016]	[1.013]	[0.963]	[1.007]	[0.981]	[0.972]	[1.006]	[1.012]
β_2	5	6.070	5.165	5.075	5.649	5.168	5.036	5.567	5.227	5.095
		(8.127)	(0.734)	(0.543)	(2.504)	(0.823)	(0.532)	(1.607)	(0.871)	(0.559)
		[5.229]	[5.060]	[5.018]	[5.109]	[5.099]	[4.972]	[5.224]	[5.089]	[5.029]
σ_2	1	1.149	0.860	0.884	0.940	0.885	0.884	0.931	0.849	0.882
		(3.677)	(0.623)	(0.507)	(1.220)	(0.655)	(0.527)	(0.968)	(0.646)	(0.497)
		[0.779]	[0.894]	[0.950]	[0.652]	[0.897]	[0.949]	[0.755]	[0.842]	[0.917]
R=500 Halton Draws										
β_1	2	2.282	2.063	2.037	2.226	2.066	2.018	2.184	2.104	2.043
		(1.884)	(0.309)	(0.224)	(1.042)	(0.343)	(0.228)	(0.629)	(0.365)	(0.240)
		[2.047]	[2.019]	[2.020]	[1.997]	[2.042]	[2.000]	[2.049]	[2.050]	[2.022]
σ_1	1	1.069	1.013	1.016	1.036	1.008	0.976	0.998	1.024	1.021
		(1.607)	(0.375)	(0.275)	(0.979)	(0.413)	(0.250)	(0.676)	(0.412)	(0.273)
		[0.930]	[1.018]	[1.015]	[0.951]	[0.995]	[0.975]	[0.961]	[1.003]	[1.017]
β_2	5	5.778	5.151	5.074	5.625	5.148	5.030	5.532	5.222	5.107
		(4.999)	(0.742)	(0.538)	(2.596)	(0.818)	(0.530)	(1.525)	(0.868)	(0.563)
		[5.182]	[5.038]	[5.013]	[5.086]	[5.075]	[4.983]	[5.199]	[5.068]	[5.052]
σ_2	1	0.955	0.808	0.856	0.882	0.836	0.855	0.888	0.816	0.877
		(1.966)	(0.660)	(0.536)	(1.267)	(0.670)	(0.548)	(0.963)	(0.673)	(0.525)
		[0.630]	[0.842]	[0.936]	[0.518]	[0.858]	[0.935]	[0.700]	[0.860]	[0.957]

[a] Mean parameter value across 500 Monte Carlo simulations.
[b] Numbers in parentheses are standard deviations of parameters across 500 Monte Carlo simulations.
[c] Numbers in brackets are median parameters across 500 Monte Carlo simulations.

Table IV-3. Results of 500 Monte Carlo Experiments: High Preference Heterogeneity

		SAS			NLOGIT			STATA		
Parameter	True Value	N=200	N=500	N=1000	N=200	N=500	N=1000	N=200	N=500	N=1000
R=100 Halton Draws										
β_1	2	2.905[a]	2.220	2.075	2.335	2.197	2.045	2.433	2.280	2.023
		(2.729)[b]	(0.861)	(0.423)	(1.215)	(0.821)	(0.422)	(2.025)	(1.439)	(0.431)
		[2.030][c]	[2.043]	[2.015]	[2.037]	[2.037]	[1.990]	[1.959]	[2.077]	[1.979]
σ_1	2	3.109	2.259	2.083	2.314	2.225	2.025	2.414	2.365	2.008
		(3.825)	(1.188)	(0.601)	(1.697)	(1.033)	(0.519)	(2.435)	(2.034)	(0.578)
		[1.964]	[2.044]	[2.032]	[1.938]	[2.049]	[1.993]	[1.954]	[2.086]	[1.937]
β_2	5	7.173	5.536	5.199	5.923	5.410	5.089	6.200	5.642	5.109
		(6.035)	(2.184)	(0.993)	(3.094)	(1.882)	(0.982)	(4.784)	(3.436)	(0.998)
		[5.188]	[5.099]	[5.040]	[5.140]	[5.031]	[4.930]	[5.137]	[5.221]	[4.945]
σ_2	5	7.214	5.562	5.223	5.933	5.447	5.096	6.166	5.706	5.121
		(6.279)	(2.310)	(1.094)	(3.284)	(2.025)	(1.037)	(4.618)	(3.731)	(1.075)
		[5.186]	[5.069]	[5.076]	[5.093]	[5.049]	[4.953]	[5.144]	[5.150]	[4.990]
R=500 Halton Draws										
β_1	2	2.729	2.198	2.099	2.344	2.182	2.066	2.380	2.266	2.040
		(2.332)	(0.718)	(0.431)	(1.303)	(0.708)	(0.426)	(1.470)	(1.270)	(0.430)
		[2.060]	[2.053]	[2.038]	[2.002]	[2.051]	[1.998]	[2.001]	[2.084]	[1.988]
σ_1	2	2.846	2.227	2.111	2.318	2.209	2.051	2.381	2.343	2.027
		(3.210)	(0.998)	(0.603)	(1.743)	(0.924)	(0.524)	(1.947)	(1.748)	(0.575)
		[1.969]	[2.057]	[2.059]	[1.920]	[2.032]	[2.007]	[1.997]	[2.109]	[1.967]
β_2	5	6.724	5.475	5.249	5.894	5.378	5.139	6.018	5.607	5.149
		(4.977)	(1.756)	(1.015)	(3.042)	(1.703)	(1.008)	(3.270)	(2.867)	(1.008)
		[5.160]	[5.061]	[5.094]	[5.086]	[5.053]	[4.992]	[5.147]	[5.227]	[5.014]
σ_2	5	6.726	5.498	5.269	5.922	5.411	5.145	5.996	5.668	5.155
		(5.070)	(1.881)	(1.114)	(3.335)	(1.843)	(1.061)	(3.450)	(3.181)	(1.079)
		[5.243]	[5.081]	[5.115]	[5.187]	[5.082]	[4.974]	[5.114]	[5.201]	[5.007]

[a] Mean parameter value across 500 Monte Carlo simulations.
[b] Numbers in parentheses are standard deviations of parameters across 500 Monte Carlo simulations.
[c] Numbers in brackets are median parameters across 500 Monte Carlo simulations.

Table IV-4. Mean Root Mean Square Errors (RMSEs) by Software Package, Number of Replications (R), and Sample Sizes (N)

	R=100 draws			R=500 draws		
True Value	N=200	N=500	N=1000	N=200	N=500	N=1000
SAS						
(2:0, 5:0)	18.449[a]	2.068	1.477	**3.102**	1.991	1.420
(2:1, 5:1)	16.737	**2.077**	1.567	10.528	**2.133**	1.599
(2:2, 5:5)	19.912	6.724	3.163	16.372	5.527	3.244
Total	55.098	10.869	6.207	30.002	9.651	6.263
NLOGIT						
(2:0, 5:0)	**3.200**	**1.916**	**1.442**	3.809	**1.875**	**1.409**
(2:1, 5:1)	5.709	2.261	**1.552**	5.982	2.282	**1.576**
(2:2, 5:5)	**9.620**	**5.896**	**2.969**	**9.742**	**5.306**	**3.044**
Total	18.529	10.073	5.963	19.533	9.463	6.029
STATA						
(2:0, 5:0)	5.731	2.139	1.459	3.356	2.072	1.433
(2:1, 5:1)	**4.041**	2.353	1.592	**3.912**	2.383	1.628
(2:2, 5:5)	14.222	10.814	3.093	10.508	9.260	3.114
Total	23.994	15.306	6.144	17.776	13.715	6.175

[a] Numbers are the sum of mean RMSEs of four estimated parameters relative to the true values.

Table IV-5. F-Statistics from ANOVA Tests for Effects of Number of Sample, Number of Replications, and Software Packages on RMSE

Variable	Degrees of Freedom	F-Statistics
Sample size	2	25.00*
Number of replications	1	2.11
Software package	2	3.53**
Sample size × number of replications	2	1.56
Sample size × software package	4	4.24*
Number of replications × software package	2	1.05
R^2		0.67
Number of observations		54

Note: One (*) and two (**) asterisks denote significance at the 1% and 5% levels of statistical significance, respectively.

Table IV-6. Default Options Employed by Three Software Packages

	Software Package		
Option	SAS	NLOGIT	STATA
Estimation method	Simulated ML	Simulated ML	Simulated ML
Optimization algorithm	Quasi-Newton	BFGS	Newton-Raphson
Convergence criteria			
Gradient	1e-5	1e-5	1e-5
Parameter	0	0	1e-4
Function	0	0	0
Initial starting value			
Mean	?	MNL estimates	MNL estimates
Standard Deviation	?	0	0.1

Note: For all software packages, Halton draws were used to simulate the likelihood function. The number of draws was either 100 or 500 depending on the simulation scenario.

Table IV-7. Run-Times in Minutes of 500 Monte Carlo Simulations by Software Packages, Number of Replications(R), and Sample Size(N)

	R=100 draws			R=500 draws		
True Value	N=200	N=500	N=1000	N=200	N=500	N=1000
SAS						
(2:0, 5:0)	15:42	40:11	77:39	74:29	188:19	377:16
(2:1, 5:1)	18:21	45:52	89:40	92:34	240:31	478:46
(2:2, 5:5)	21:02	42:15	81:53	102:41	217:16	427:46
NLOGIT						
(2:0, 5:0)	55:44	137:50	274:46	259:13	670:43	1376:08
(2:1, 5:1)	73:16	178:02	354:36	382:15	889:13	1801:56
(2:2, 5:5)	141:13	337:11	2050:32	741:05	1858:20	3812:53
STATA						
(2:0, 5:0)	44:42	108:21	214:10	156:08	388:21	787:39
(2:1, 5:1)	64:27	137:27	275:15	231:51	507:36	1022:06
(2:2, 5:5)	228:32	235:46	343:00	686:09	875:04	1247:20

APPENDIX IV-A

Table IV-A. 1. Table Size of the Test: Percent of Cases when Significant Taste Heterogeneity was Incorrectly Observed (NLOGIT and *R*=100)

	σ_1		σ_2		σ_1 and σ_2	
Sample Size	α = 5% or lower	α = 10% or lower	α = 5% or lower	α = 10% or lower	α = 5% or lower	α = 10% or lower
N=200	2.8%	6.8%	3.0%	7.0%	0.4%	0.4%
N=500	5.2%	9.4%	7.6%	10.6%	0.4%	0.4%
N=1000	8.2%	11.6%	8.6%	12.6%	0.8%	1.0%

Table IV-A. 2. Comparison of RPL Estimates by Optimization Algorithm and Software Package: *N*=200 and *R*=100 case

Parameter	True Value	SAS QN*	SAS NR	NLOGIT BFGS*	NLOGIT NR	STATA NR*	STATA BFGS
β_1	2	2.386[a]	2.312	2.231	2.231	2.197	2.288
		(3.034)[b]	(1.975)	(0.990)	(0.994)	(0.652)	(2.020)
		[2.073][c]	[2.084]	[2.000]	[2.000]	[2.054]	[2.062]
σ_1	1	1.113	1.102	1.050	1.050	1.021	1.158
		(1.814)	(1.606)	(0.889)	(0.892)	(0.689)	(2.987)
		[0.939]	[0.936]	[0.963]	[0.963]	[0.972]	[0.980]
β_2	5	6.070	5.869	5.649	5.650	5.567	5.849
		(8.127)	(5.286)	(2.504)	(2.510)	(1.607)	(6.180)
		[5.229]	[5.202]	[5.109]	[5.109]	[5.224]	[5.235]
σ_2	1	1.149	1.071	0.940	0.940	0.931	0.986
		(3.677)	(2.234)	(1.220)	(1.220)	(0.968)	(1.453)
		[0.779]	[0.748]	[0.652]	[0.652]	[0.755]	[0.743]

Note: Algorithms QN, NR, and BFGS represent Quasi Newton, Newton-Raphson, and Broyden, Fletcher, Goldfarb, Shanno methods, respectively.

A single asterisk (*) indicates a default algorithm for each software package.

[a] Mean parameter value across 500 Monte Carlo simulations.

[b] Numbers in parentheses are standard deviations of parameters across 500 Monte Carlo simulations.

[c] Numbers in brackets are median parameters across 500 Monte Carlo simulations.

Table IV-A. 3. Comparison of RPL Estimates by Random Numbers obtained with and Loaded into Software Packages: *N*=200 and *R*=100 Case

Parameter	True Value	SAS - SAS	SAS - NLOGIT	LIMDEP - NLOGIT
β_1	2	2.386[a]	2.234	2.231
		(3.034)[b]	(0.897)	(0.990)
		[2.073][c]	[2.079]	[2.000]
σ_1	1	1.113	1.050	1.050
		(1.814)	(0.863)	(0.889)
		[0.939]	[0.961]	[0.963]
β_2	5	6.070	5.669	5.649
		(8.127)	(2.283)	(2.504)
		[5.229]	[5.237]	[5.109]
σ_2	1	1.149	1.015	0.940
		(3.677)	(1.183)	(1.220)
		[0.779]	[0.825]	[0.652]

[a] Mean parameter value across 500 Monte Carlo simulations.
[b] Numbers in parentheses are standard deviations of parameters across 500 Monte Carlo simulations.
[c] Numbers in brackets are median parameters across 500 Monte Carlo simulations.

Appendix IV-B: Examples of the Software Program Code

IV-B.1. SAS

```
dm 'log;clear;output;clear;';

* Monte Carlo Experiment for testing RPL Model
* Sample size 200_5

data raw200;
ss=57139;
do sample=1 to 500;
do obs=1 to 200;
x1=rannor(ss); x11=rannor(ss);
x2=rannor(ss); x22=rannor(ss);
x3=rannor(ss); x33=rannor(ss);
b1=rannor(ss); b2=rannor(ss);
h1=ranuni(ss); h2=ranuni(ss); h3=ranuni(ss);
e1=-log(log(1/h1)); e2=-log(log(1/h2)); e3=-log(log(1/h3));

a1=2; a2=5; d1=2; d2=5;

u1=(a1+b1*d1)*x1+(a2+b2*d2)*x11+e1;
u2=(a1+b1*d1)*x2+(a2+b2*d2)*x22+e2;
u3=(a1+b1*d1)*x3+(a2+b2*d2)*x33+e3;

if u1>u2 and u1>u3 then c1=1; else c1=0;
if u2>u1 and u2>u3 then c2=1; else c2=0;
if u3>u1 and u3>u2 then c3=1; else c3=0;
output;
end; end;

proc sort; by sample;
proc means data=raw200; by sample;

data orig; set raw200;
keep x1 x11 x2 x22 x3 x33 c1 c2 c3 u1 u2 u3 sample;
run;

data new (keep=sample id mode x xx u c) ;  set orig;
array var1{3} x1 x2 x3;
array var2{3} x11 x22 x33;
array util{3} u1 - u3;
array ch{3} c1 - c3;
retain id 0;
id+1;
do i=1 to 3;
mode=i;
x=var1{i};
xx=var2{i};
u=util{i};
c=ch{i};
output;
end;
run;

data mc200; set new;
proc mdc data=mc200 outest=out1 ; by sample;
model c= x xx / nchoice = 3 type=mxl nsimul=500
mixed=(normalparm=x xx) ;
id id ;
run;

proc means data=out1 n mean p50 std p5 p95;
run;
quit;
```

IV-B.2. NLOGIT

```
/* Monte Carlo Method: Sample Size - 200_5  */
reset
timer
calc; ran(57139) $

proc=mcset $
calc; ni=500; i=0 $
sample; 1-200 $
matrix; estb=init(ni,4,0) $
endproc

proc=mcrun $
calc; i=0 $
label; 9999 $
sample; 1-200 $
create; a1=2; a2=5; d1=2; d2=5 $
create; h1=rnu(0,1); h2=rnu(0,1); h3=rnu(0,1) $
create; e1=-log(log(1/h1)); e2=-log(log(1/h2)); e3=-log(log(1/h3)) $
create; b1=rnn(0,1); b2=rnn(0,1) $
create; x1=rnn(0,1); x11=rnn(0,1); x2=rnn(0,1); x22=rnn(0,1); x3=rnn(0,1); x33=rnn(0,1) $
create; u1=(a1+d1*b1)*x1+(a2+d2*b2)*x11+e1;
        u2=(a1+d1*b1)*x2+(a2+d2*b2)*x22+e2;
        u3=(a1+d1*b1)*x3+(a2+d2*b2)*x33+e3 $
create; if (u1>u2 & u1>u3) c1=1; (else) c1=0;
        if (u2>u1 & u2>u3) c2=1; (else) c2=0;
        if (u3>u1 & u3>u2) c3=1; (else) c3=0 $
create; if (c1=1) choice=1; if (c2=1) choice=2; if (c3=1) choice=3 $


rplogit; lhs=choice
       ; choices=c1, c2, c3 [1]
       ; rhs=x1,x2,x3,x11,x22,x33
       ; attr=x,xx
       ; fcn=x(n),xx(n)
       ; halton ; pds=1; maxit=100; pts=500 $
matrix; estb(i,*)=b $

calc; list; i=i+1 $
go to; 9999; i<=500 $

endproc

exec; proc=mcset $
exec; proc=mcrun $

sample; 1-500 $
matrix; totest=estb $
matrix; xm=part(totest,1,500,1,1); xxm=part(totest,1,500,2,2);
        xs=part(totest,1,500,3,3); xxs=part(totest,1,500,4,4) $
create; mcxm=xm $
create; mcxs=xs $
create; mcxxm=xxm $
create; mcxxs=xxs $
dstat; rhs=mcxm, mcxs, mcxxm, mcxxs; quantiles $
stop $
```

IV-B.3. STATA

```
* Monte Carlo Simulation: 200_5
clear
set seed 57139
set more off
set mem 100m
set matsize 500
local B=500
matrix xm=J(`B',1,0)
matrix xxm=J(`B',1,0)
matrix xs=J(`B',1,0)
matrix xxs=J(`B',1,0)
timer clear 1

forvalues b=1/`B' {
timer on 1
drop _all
quietly set obs 200
gen id=_n
gen x1=invnorm(uniform())
gen xx1=invnorm(uniform())
gen x2=invnorm(uniform())
gen xx2=invnorm(uniform())
gen x3=invnorm(uniform())
gen xx3=invnorm(uniform())
gen b1=invnorm(uniform())
gen b2=invnorm(uniform())
gen h1=uniform()
gen h2=uniform()
gen h3=uniform()
gen a1=2
gen a2=5
gen d1=2
gen d2=5
gen e1=-log(log(1/h1))
gen e2=-log(log(1/h2))
gen e3=-log(log(1/h3))
gen u1=(a1+d1*b1)*x1+(a2+d2*b2)*xx1+e1
gen u2=(a1+d1*b1)*x2+(a2+d2*b2)*xx2+e2
gen u3=(a1+d1*b1)*x3+(a2+d2*b2)*xx3+e3

gen choice=.
replace choice=1 if u1>u2 & u1>u3
replace choice=2 if u2>u1 & u2>u3
replace choice=3 if u3>u1 & u3>u2

case2alt, alt(x xx) case(id) choice(choice) gen(choice2) altnum(mode)

mixlogit choice2, group(id) rand(x xx) nrep(500) iterate(100)
matrix beta=e(b)

matrix xm[`b',1]=beta[1,1]
matrix xxm[`b',1]=beta[1,2]
matrix xs[`b',1]=beta[1,3]
matrix xxs[`b',1]=beta[1,4]

timer off 1
}
timer list 1
drop _all
svmat xm
svmat xxm
svmat xs
svmat xxs

summ *, det
```

CHAPTER V

CONCLUSIONS

Stated preference data obtained through real and hypothetical surveys and experiments are frequently used to determine consumers demand for product attributes. Although such data are increasingly being used in economic research, there remain doubts about the validity of preference elicitation methods and econometric models used to estimate consumer preferences. This study explores such doubts and provides richer understanding of consumer demand.

The first essay addressed an important question in experimental economics, stated preference methods, and conjoint analysis: whether elicited values are consisted with observed values in the field. This study investigated the ability of three preference elicitation methods (hypothetical choices, non-hypothetical choices, and non-hypothetical rankings) and three discrete choice econometric models (the MNL, the IAL, and RPL) to predict actual retail shopping behavior for twelve products in three different product categories (ground beef, wheat flour, and dishwashing liquid). Participants were randomly assigned to one of three experimental treatments. Sales data provided by the local grocery store was used to calculate the true field market share of each good in each product category. This study confirmed the implicit assumption made in much experimental work: that non-hypothetical lab valuations more closely mirror field behavior than hypothetical valuations. Results indicated that all elicitation methods exhibited a high level of external validity. Overall, non-hypothetical approaches, especially the non-hypothetical ranking method, predicted retail sales more accurately than the hypothetical choice experiment. The RPL model, which relaxes the assumption of the independence of irrelevant alternatives of the MNL by modeling preference heterogeneity, exhibited the best predictive performance followed by the MNL. However, the IAL model, which relaxes the assumption of deterministic choice sets of the MNL, never outperformed the MNL and RPL models.

The second study incorporated fairness concerns into a field: food consumption. Although a number of experimental studies have found other-regarding behavior of participants in their simulated settings, few studies have actually discovered such other-regarding behavior in the field. This essay determined whether fairness concerns carry over to food choice and investigated the extent to which the fairness models proposed in the general economics literatures explain food purchasing behavior. Mail surveys were developed and sent to a random sample of consumers. Responds indicated the likelihood of purchasing a loaf of bread which had a particular price and a specified amount of profit from the purchase going to the participants in the food supply chain: small farmers, large farmers, agribusiness processors, and grocery stores.

Although none of all fairness concerns considered in this essay were identified as statistically significant, or even one (ERC) was significant, the opposite sign was found than expected sign, all models suggested that consumers do care about the benefits to small farmers. Moreover, people believed every participant, especially small farmers, in the food marketing channel benefit more from selling organic bread than conventional non-organic bread. These results might explain high portion (from 39.9% to 48.8%) of premium for organic bread over conventional bread which represents consumers' concerns for the distribution of profits across participants in the supply chain. In general, however, results indicated that the general fairness models proposed in economic literatures have little explanatory power in the food purchasing decision. With concerns about unclearness of perception of consumers' own payouts, original fairness models were modified by excluding profits to self ("consumers"). All modified models revealed better out-of-sample predictive performance by two criteria, the MSE and OSLLF. All estimates for fairness concerns were statistically significant and expected sign. For example, the fairness coefficient in the FS model was positive (5.727), implying that people dislike the payoff difference between small farmers and any other participants in the food supply chain. As previously watched, a premium for organic bread described by concerns for fairness accounted for about 40% of total willingness-to-pay premium.

The third paper used Monte Carlo simulations to determine the sensitivity of RPL estimates. As a result of conceptual advantages and realistic assumption imposed, the RPL model has been increasingly used in various research areas, resulting in the further development of econometric software packages which provide user-friendly "canned" routines to estimate the RPP models. In this study, the sensitivity of RPL estimates was determined with a variety of factors: the level of preference heterogeneity, sample size, number of replications used in the simulated maximum likelihood approach, and econometric software package.

Results from Monte Carlo experiments indicated RPL estimates exhibit bias in small sample and this tendency diminished by increasing sample size. However, in large sample, it was observed high possibility of type I error. The standard MNL model fits better than the RPL with simulated data which has no taste heterogeneity. While, surprisingly, the number of replications in simulated maximum likelihood function did not have much influence on the RPL estimates, there was difference between the software packages. In general, LIMDEP-NLOGIT generated most efficient estimates in terms of RMSE.

Results of this dissertation not only enhance our understanding of the preference elicitation methods and econometric models frequently employed in studies of consumer demand but also provoke several issues for additional researches to improve the current study. In the first essay, three preference elicitation methods were utilized to compare the external validity of surveys and experiments. Although these three methods are widely used, there are alternative approaches, such as experimental auction auctions. Recently, this method has captured much attention of researchers to determine consumer behavior for several advantages such as incentive compatibility and flexibility to use. In addition to preference elicitation methods, several econometric models - for example, latent class model, nested logit model, and generalized extreme value model - can be also compared to predict actual shopping behavior. Including more methods and models and, if possible, providing additional results for high levels of external validity of those should increase confidence for the elicited values.

As found in the first essay, hypothetical elicitation approaches often produce overestimated preference values, a phenomenon referred to as hypothetical bias. In the second study, the hypothetical mail survey method was used, thus there is possibility of distortion in the results. Conducting an experimental method with real monetary payments and comparing this new study to the second essay might be interesting for further studies. While the "fairness" motive was incorporated into consumers' food purchasing behavior in the second essay, this motive can be also applied to other applications. For example, consumers might have preference for the government subsidies to farms or the direct payment programs to defend domestic farmers associated with international trade, which might be explained by fairness concerns for farmers or agricultural sector.

Results in the third essay show differences in RPL performance across software packages in small samples. Although several possible explanations were addressed, further research is needed to examine in detail why large differences in RPL estimates exist across software packages. Moreover, this study can be expanded by considering welfare values (e.g., willingness-to-pay) calculated by using estimates. In estimating the RPL model, practitioners should assume the distribution of the random parameters. It means that different distribution assumption for the random parameters may yield different parameter estimates, as results, different preference values.

Future studies addressed the aforementioned concerns are expected to provide researchers, agribusiness, and policy makers with more in-depth insights on consumer behavior.

REFERENCES

Adamowicz, W., R. Boxall, M. Williams, and J. Louviere. 1998. "Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation." *American Journal of Agricultural Economics* 80:64-75.

Adamowicz, W., J. Louviere, and M. Williams. 1994. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26:271-92.

Alfnes, F. A.G. Guttormsen, G. Steine, and K. Kolstad. 2006. "Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives." *American Journal of Agricultural Economics* 88:1050-61.

Andreoni, J., and J. Miller. 2002. "Giving According to GARP: An Experimental Test of the Rationality of Altruism." *Econometrica* 70:737-53.

Andrews, R.L., and T.C. Srinivasan. 1995. "Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data." *Journal of Marketing Research* 32:30-41.

Baiocchi, G. 2005. "Monte Carlo Methods in Environmental Economics." In R. Scarpa and A. Alberini eds. *Applications of Simulation Methods in Environmental and Resource Economics*. The Netherlands: Springer, pp.317-40.

Bajari, P., J.T. Fox, and S.P. Ryan. 2007. "Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients." *American Economic Review* 97:459-63.

Beggs, S., S. Cardell, and J. Hausman. 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics* 17:1-19.

Ben-Akiva, M., and B. Boccara. 1995. "Discrete Choice Models with Latent Choice Sets." *International Journal of Research in Marketing* 12:9-24.

Bereby-Meyer, Y., and M. Niederle. 2005. "Fairness in Bargaining." *Journal of Economic Behavior & Organization* 56:173-86.

Bhat, C.R. 2001. "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." *Transportation Research Part B* 35:677-693.

Bockstael, N.E., and K.E. McConnell. 2007. *Environmental and Resource Valuation with Revealed Preferences*. The Netherlands: Springer.

Boxall, P.C., J. Englin, and W.L. Adamowicz. 2003. "Valuing aboriginal artifacts: A Combined Revealed-Stated Preference Approach." *Journal of Environmental Economics and Management* 45:213-30.

Bolton, G.E. and A. Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90:166-93.

______. 2006. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment." *American Economic Review* 96:1906-11.

Boyle, K.J., T.P. Holmes, M.F. Teisl and B. Roe. 2001. "A Comparison of Conjoint Analysis Response Formats." *American Journal of Agricultural Economics* 83:441-54.

Brookshire, D.S., D.L. Coursey, and W.D. Schulze. 1987. "The External Validity of Experimental Economics Techniques: Analysis of Demand Behavior." *Economic Inquiry* 25:239-50.

Brookshire, D.S., M. Thayer, W.D. Schulze, R. d'Arge. 1982. "Valuing Public Goods: A Comparison of the Survey and Hedonic Approaches." *American Economic Review* 72:165-77.

Brownstone, D., D. Bunch, and K. Train. 2000. "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles." *Transportation Research B* 34:315-38.

Carson, R.T., N.E. Flores, K. Martin, J.L. Wright. 1996. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 71:80-99.

Carson, R.T., and T. Groves. 2007. "Incentive and informational properties of preference questions." *Environmental and Resource Economics* 37:181-210.

Charness, G., and M. Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117:817-69.

Chiou, L. and J.L. Walker. 2007. "Masking identification of discrete choice models under simulation methods." *Journal of Econometrics* 141: 683-703.

Dillman, D.A. 2000. *Mail and Internet Surveys*. New York: John Wiley & Sons, Inc.

Ding, M., R. Grewal, and J. Liechty. 2005. "Incentive-Aligned Conjoint Analysis." *Journal of Marketing Research* 42:67-83.

Engelmann, D., and M. Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." *American Economic Review* 94:857-69.

Erlei, M. 2008. "Heterogeneous Social Preferences." *Journal of Economic Behavior & Organization* 65:436-57.

Fehr, E., G. Kirchsteiger, and A. Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108:437-59.

Fehr, E., and K.M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114:817-68.

Fehr, E., M. Naef, and K.M. Schmidt. 2006. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment." *American Economic Review* 96:1912-17.

Fox, J., J. Shogren, D. Hayes, and J. Kliebenstein. 1998. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics* 80:455-65.

Güth, W., R. Schmittberger, and B. Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3:367-88.

Haab, T.C., and R.L. Hicks. 1997. "Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand." *Journal of Environmental Economics and Management* 34:127-47.

Harrison, G., and J.A. List. 2004. "Field Experiments." *Journal of Economic Literatures* 42:1009-55.

Huber, J., and K. Train. 2000. "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths." *Marketing Letter* 12:259-69.

International Federation of Organic Agriculture. Available at http://www.ifoam.org, accessed 14 April 2008.

Jones, S. and D.A. Hensher. 2007. "Evaluating the Behavioral Performance of Alternative Logit Models: An Application to Corporate Takeovers Research." *Journal of Business Finance & Accounting* 34:1193-1220.

Johnston, R.J. 2006. "Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum." *Journal of Environmental Economics and Management* 52:469-81.

Kastens, T.L., and G.W. Brester. 1996. "Model Selection and Forecasting Ability of Theory-Constrained Food Demand Systems." *American Journal of Agricultural Economics* 78:301-12.

Krinsky, I., and A.L. Robb. 1986. "On Approximating the Statistical Properties of Elasticities." *The Review of Economics and Statistics* 64:715-19.

Levitt, S.D., and J.A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives* 21:153-74.

List, J.A. 2006. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effect in Actual Transactions." *Journal of Political Economy* 114: 1-37.

List, J.A., and C.A. Gallet. 2001. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20: 241-54.

Little, J., and R. Berrens. 2004. "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis." *Economics Bulletin* 3:1-13.

Loomis, J., M. Creel., and T. Park. 1991. “Comparing Benefit Estimates from Travel Cost and Contingent Valuation Using Confidence Intervals for Hicksian Welfare Measures.” *Applied Economics* 23:1725-31.

Louviere, J.J., D.A. Hensher, and J.D. Swait. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge, UK: Cambridge University Press.

Lusk, J.L., D. Fields, and W. Prevatt. 2008. “An Incentive Compatible Conjoint Ranking Mechanism.” *American Journal of Agricultural Economics* 90:487-98.

Lusk, J.L., B. Norwood, and R. Pruitt. 2006. “Consumer Demand for a Ban on Subtherapeutic Antibiotic Drug Use in Pork Production.” *American Journal of Agricultural Economics* 88:1015-1033.

Lusk, J.L., J.R. Pruitt, and F.B. Norwood. 2006. “External Validity of a Framed Field Experiment.” *Economic Letters* 93: 285-90.

Lusk, J.L., J. Roosen, and J.A. Fox. 2003. “Demand for Beef from Cattle Administered Growth Hormones or Fed Genetically Modified Corn: A Comparison of Consumers in France, Germany, the United Kingdom, and the United States.” *American Journal of Agricultural Economics* 85:16-29.

Lusk, J.L., and T.C. Schroeder. 2004. “Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks.” *American Journal of Agricultural Economics* 86: 467-82.

Manski, C.F. 1997. “The Structure of Random Utility Models.” *Theory and Decision* 8:229-54.

McFadden, D. 1974. “Conditional Logit Analysis of Qualitative Choice Behavior.” In P. Zarembka ed., *Frontiers in Econometrics* New York: Academic Press, pp.105-42.

McFadden, D., and K.Train, 2000. “Mixed MNL Models for Discrete Response.” *Journal of Applied Econometrics* 15:447-70.

McCullough, B.D., and H.D. Vinod. 1999. “The Numerical Reliability of Econometric Software.” *Journal of Economic Literature* 37:633-65.

Murphy, J.J., P.G.. Allen, T.H. Stevens, and D. Weatherhead. 2005. “A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation.” *Environmental and Resource Economics* 30:313–325.

Murphy, E., B. Norwood, and M. Wohlgenant. 2004. “Do Economic Restrictions Improve Forecasts?” *Journal of Agricultural and Applied Economics* 36:549-58.

Narayanan, S., R. Desiraju, and P.K. Chintagunta. 2004. “Return on Investment Implications for Pharmaceutical Promotional Expenditures: The Role of Marketing-Mix Interactions.” *Journal of Marketing* 68:90-105.

Norwood, F. B., J. L. Lusk, and B. W. Brorsen. 2004. “Model Selection for Discrete Dependent Variables: Better Statistics for Better Steaks.” *Journal of Agricultural and Resource Economics* 29:404-19.

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

Revelt, D., and K. Train. 1998. “Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level.” *Review of Economics and Statistics* 80: 647-57.

Shogren, J. F., J. A. Fox, D. J. Hayes, and J. Roosen. 1999. “Observed for Food Safety in Retail, Survey, and Auction Markets.” *American Journal of Agricultural Economics* 81:1192-9.

Swait, J., and M. Ben-Akiva. 1987. “Incorporating Random Constraints in Discrete Models of Choice Set Generation.” *Transportation Research B* 21B:91-102.

Swait, J., and J. Louviere. 1993. “The Role of the Scale Parameter in the Estimation and Use of Multinomial Logit Models.” *Journal of Marketing Research* 30: 305-14.

Train, K.E. 1999. “Halton Sequences for Mixed Logit.” Available at http://elsa.berkeley.edu/~train/papers.html.

_______. 2001. “A Comparision of Hierarchical Bayes and Maximum Simulated Likelihoood for Mixed Logit.” Available at http://elsa.berkeley.edu/~train/papers.html.

_______. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Train, K., and M. Weeks. 2005. “Discrete Choice Models in Preference Space and Willingness-to-Pay Space.” In R. Scarpa and A. Alberini eds. *Applications of Simulation Methods in Environmental and Resource Economics.* Netherlands: Springer, pp. 1-16.

Vossler, C.A., J. Kerkvliet, S. Polasky, and O. Gainutdinova. 2003. “Externally Validating Contingent Valuation: An Open Space Survey and Referendum in Corvallis, Oregon.” *Journal of Economic Behavior and Organization* 51:261-77.

Vossler, C.A., and J. Kerkvliet. 2003. “A Criterion Validity Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior for a

Public Referendum." *Journal of Environmental Economics and Management* 45:631-49.

APPENDICES

APPENDIX A. INSTITUTIONAL REVIEW BOARD LETTER

Oklahoma State University Institutional Review Board

Date:	Wednesday, March 28, 2007
IRB Application No	AG0715
Proposal Title:	Consumer Preferences for Food Production Practices
Reviewed and Processed as:	Exempt

Status Recommended by Reviewer(s): Approved Protocol Expires: 3/27/2008

Principal
Investigator(s

Jayson Lusk
411 Ag Hall
Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☑ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Sue C Jacobs

Sue C. Jacobs, Chair
Institutional Review Board

APPENDIX B. SURVEY ON CONSUMER PREFERENCES FOR FOOD PRODUCTION PRACTICES

Oklahoma State University

Study on

Consumer Preferences
for
Food Production Practices

Consumer Preferences for Food Production Practices

1. **Are you more or less concerned about the way most food is produced today as compared to five years ago?**
   - ❑ Less concerned today than 5 years ago
   - ❑ About the same level of concern
   - ❑ More concerned today than 5 years ago

2. **Have you ever purchased organic food?** ***(Organic food refers to food produced under a production system designed to promote biodiversity and soil biological activity. Organic food is typically grown without synthetic pesticides, fertilizers, or genetically modified ingredients.)***
   - ❑ Yes
   - ❑ No
   - ❑ I don't know

3. **What is the largest premium you would be willing to pay for an organic loaf of bread over a loaf of bread produced through conventional, non-organic means, assuming both loaves were the same brand name?**
   - ❑ I am not willing to pay a premium for organic bread
   - ❑ $0.01 to $0.49 premium
   - ❑ $0.50 to $0.99 premium
   - ❑ $1.00 to $1.49 premium
   - ❑ $1.50 to $1.99 premium
   - ❑ $2.00 premium or more

4. **How much do you think <u>small farmers</u> (farming less than 500 acres) profit from the sale of a single loaf of bread?** ***(please check a box for each of the following production practices)***

	$0.01 to $0.05	**$0.06 to $0.10**	**$0.11 to $0.15**
Organic bread	❑	❑	❑
Conventional non-organic bread	❑	❑	❑

5. **How much do you think <u>large farmers</u> (farming more than 500 acres) profit from the sale of a single loaf of bread?** ***(check a box for each production practice)***

	$0.01 to $0.05	**$0.06 to $0.10**	**$0.11 to $0.15**
Organic bread	❑	❑	❑
Conventional non-organic bread	❑	❑	❑

6. **How much do you think <u>agribusiness processors</u> (wheat millers and bakers) profit from the sale of a single loaf of bread?** ***(check a box for each production practice)***

	$0.01 to $0.05	**$0.06 to $0.10**	**$0.11 to $0.15**
Organic bread	❑	❑	❑
Conventional non-organic bread	❑	❑	❑

7. **How much do you think your <u>local grocery store chain</u> profits from the sale of a single loaf of bread?** ***(check a box for each production practice)***

	$0.01 to $0.05	**$0.06 to $0.10**	**$0.11 to $0.15**
Organic bread	❑	❑	❑
Conventional non-organic bread	❑	❑	❑

8. **What is the most you would be willing to pay for a loaf of bread if no other market participant (farmers, agribusinesses, or grocery store) made any profit and you and your family were the only people that benefited from the purchase?** ***(please write the most you are willing to pay for each production practice in the in the blanks below, assuming you and your family were the only people to benefit from the purchase)***

The most I would be willing to pay for organic bread is . . . . $_____ per loaf

The most I would be willing to pay for conventional-non organic bread is $_____ per loaf

9. **How much would you normally expect to pay for a single loaf of bread in your local grocery store?** ***(please write the price you expect would be charged for each type of break in the blanks below)***

I would expect the price of organic bread to be . . . . $_____ per loaf

I would expect the price of conventional, non-organic bread to be . . . . $_____ per loaf

10. **To what extent do you agree or disagree with the following statement, "Organic food is safer to eat than conventional, non-organic food?"**

- ☐ Strongly agree
- ☐ Agree
- ☐ Moderately agree
- ☐ Neither agree nor disagree
- ☐ Moderately disagree
- ☐ Disagree
- ☐ Strongly disagree

11. **To what extent do you agree or disagree with the following statement, "Organic food is higher quality than conventional, non-organic food?"**

- ☐ Strongly agree
- ☐ Agree
- ☐ Moderately agree
- ☐ Neither agree nor disagree
- ☐ Moderately disagree
- ☐ Disagree
- ☐ Strongly disagree

12. **To what extent do you agree or disagree with the following statement, "Purchasing organic food helps improve the environment?"**

- ☐ Strongly agree
- ☐ Agree
- ☐ Moderately agree
- ☐ Neither agree nor disagree
- ☐ Moderately disagree
- ☐ Disagree
- ☐ Strongly disagree

13. **How likely are you to buy a loaf of bread from your local grocery store if it was produced from a food production system with the following characteristics?** ***In the first case below, the question is: how likely are you to purchase a loaf of bread if the price you would pay is $2.99, and from that single purchase small farmers would earn a profit of $0.01, large farmers would earn a profit of $0.15, agribusinesses would earn a profit of $0.01, and your local grocery store would earn a profit of $0.15? (Note: small farmers are defined here as those that farm less than 500 acres, large farmers are defined here as those that farm more than 500 acres, and agribusinesses represent wheat millers and bakers.) For each of the 6 products listed below, please circle the appropriate number to indicate your answer. Note: "0" means you definitely would NOT purchase the loaf of bread, a "5" means there is a 50/50 chance you would purchase the loaf of bread, and a "10" means you definitely would buy the loaf of bread.***

Product	Definitely Would Not Buy					Equal Chance of Buying and Not Buying					Definitely Would Buy
Price of bread loaf: **$2.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$1.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$1.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.01**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$2.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.15** Profit to grocery store: **$0.01**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$1.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.15** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$3.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.15** Profit to grocery store: **$0.01**	0	1	2	3	4	5	6	7	8	9	10

➢ **Please continue to the next page to answer 6 questions very similar to those above**

14. How likely are you to buy a loaf of bread from your local grocery store if it was produced from a food production system with the following characteristics? *Note: the structure of the questions is identical to that on the previous page. As before, please circle the appropriate number to indicate your answer for each of the 6 products shown below. A "0" means you definitely would NOT purchase the loaf of bread, a "5" means there is a 50/50 chance you would purchase the loaf of bread, and a "10" means you definitely would buy the loaf of bread.*

Product	Definitely Would Not Buy					Equal Chance of Buying and Not Buying					Definitely Would Buy
Price of bread loaf: **$1.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.15** Profit to grocery store: **$0.01**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$3.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.15** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$2.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.01**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$3.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$2.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.15** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$3.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.01**	0	1	2	3	4	5	6	7	8	9	10

➢ **Please continue to the next page of the survey**

➢ **We would like some background information about you. This is an important part of our analysis. The survey is anonymous and your name is in no way linked to the responses.**

15. What is your gender?
- ❑ Male
- ❑ Female

16. Do you or does anyone in your immediate family (grandparents, parents, siblings, aunts, or uncles) farm or ranch for a living?
- ❑ Yes
- ❑ No

17. How populated is the county in which you live?
- ❑ Fewer than 10,000 people
- ❑ Between 10,000 and 99,999 people
- ❑ Between 100,000 and 499,999 people
- ❑ More than 500,000 people

18. Is the economy of the county in which you live primarily dependent on farming?
- ❑ Yes
- ❑ No

19. Are there children under the age of 12 living in your household?
- ❑ Yes
- ❑ No

20. Have you obtained a Bachelor's degree from a university or college?
- ❑ Yes
- ❑ No

21. Have you obtained a graduate degree such as an M.S., M.A., M.B.A., Ph.D. or M.D. or J.D.?
- ❑ Yes
- ❑ No

22. What is your approximate annual household income before taxes in 2006?
- ❑ Less than $20,000
- ❑ $20,000 to $39,999
- ❑ $40,000 to $59,999
- ❑ $60,000 to $79,999
- ❑ $80,000 to $99,999
- ❑ $100,000 or more

23. What is your present age? ______ years

24. In what U.S. state do you live? ______ state

Thank you for help! Please return your survey in the pre-paid postage envelope!

VITA

Jae Bong Chang

Candidate for the Degree of

Doctor of Philosophy

Dissertation: THREE ESSAYS ON MODELING CONSUMER DEMAND

Major Field: Agricultural Economics

Biographical:

Personal Data: Born in Daegu, South Korea, on May 21, 1973, The son of Byung Joo Chang and Bok Ja Choi.

Education: Received Bachelor of Science Degree in Agricultural Economics from SungKyunKwan University, Seoul, Korea in February 1999; received Master of Science Degree in Agricultural Economics from Seoul National University, Seoul, Korea in February 2001; received Master of Science Degree in Economics from Iowa State University, Ames, IA in August 2004. Completed the requirements for the Doctor of Philosophy Degree in Agricultural Economics at Oklahoma State University, Stillwater, Oklahoma in May, 2009.

Experience: Employed as a Graduate Assistant at Seoul National University, Department of Agricultural Economics, September 1999 – December 2000; employed as a Researcher at Korea Rural Economic Institute, December 2001 – June 2002; employed as a Graduate Assistant at Iowa State University, Department of Economics, August 2002 – August 2004; employed as a Researcher at Korea Rural Economic Institute, October 2004 – July 2005; employed as a Graduate Assistant at Oklahoma State University, Department of Agricultural Economics, August 2005 – May 2009.

Professional Memberships: Agricultural and Applied Economics Association

Name: Jae Bong Chang Date of Degree: May, 2009

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THREE ESSAYS ON MODELING CONSUMER DEMAND

Pages in Study: 110 Candidate for the Degree of Doctor of Philosophy

Major Field: Agricultural Economics

Scope, Method of Study, and Findings: Although stated preference data obtained through real and hypothetical surveys and experiments are increasingly being used in economic research, there remain doubts about the validity of preference elicitation methods and econometric models used to estimate consumer preferences. This study explores such doubts and provides richer understanding of consumer demand.

The first essay compared the ability of three preference elicitation methods (hypothetical choices, nonhypothetical choices, and nonhypothetical rankings) and three discrete-choice econometric models (the multinomial logit [MNL], the independent availability logit [IAL], and the random parameter logit [RPL]) to predict actual retail shopping behavior in three different product categories (ground beef, wheat flour, and dishwashing liquid). Overall, this study found a high level of external validity. Specific results suggested that the nonhypothetical elicitation approaches, especially the nonhypothetical ranking method, outperformed the hypothetical choice experiment in predicting retail sales. This study also found that the RPL can have superior predictive performance, but that the MNL predict equally well in some circumstances.

Although experimental studies have reported a wide array of other-regarding behavior, the pervasiveness of such behavior in the field is an open question. Using a stated preference experiment, the second essay first estimated people's preferences, when purchasing food products, for the distribution of benefits accruing to participants in the food supply chain. Although none of the existing fairness models exhibit much predictive power, this study found that people are in-fact concerned about the distribution of benefits resulting from food purchases, and that modifications to the models to fit the food context significantly improves explanatory power. Finally, this study found that the measured preferences, along with elicited beliefs are significant predictors of people's willingness-to-pay a premium for organic food.

RPL or mixed logit models are increasingly being reported in the literature, but uncertainties about the reliability of the model remain. Using Monte Carlo simulations, the third paper found that mixed logit estimates exhibit bias in small samples (N=200); an effect that dissipates as sample size increases. Despite the fact that the conventional multinomial logit is a restricted form of the mixed logit, simulation results indicate that if there is no preference heterogeneity, the conventional multinomial logit provides *much* more efficient estimates than the mixed logit. Moreover, in large samples there is a high likelihood of observing type I errors – finding statistically significant heterogeneity when none exists. Finally, simulation results indicate differences in performance across commercial software package providing mixed logit estimation routines.

ADVISER'S APPROVAL: Dr. Jayson L. Lusk