



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

## **External Validity of Hypothetical Surveys and Laboratory Experiments**

**Jae Bong Chang**

Ph.D. candidate and Graduate Research Assistant  
Department of Agricultural Economics, Oklahoma State University

**Jayson L. Lusk**

Professor and Willard Sparks Endowed Chair  
Department of Agricultural Economics, Oklahoma State University

**F. Bailey Norwood**

Associate Professor  
Department of Agricultural Economics, Oklahoma State University

Contact: Jayson L. Lusk  
Department of Agricultural Economics, 411 Ag Hall  
Oklahoma State University  
Stillwater, OK 75078  
Phone: (405) 744-7465  
Fax: (405) 744-8210  
E-mail: [jayson.lusk@okstate.edu](mailto:jayson.lusk@okstate.edu)

*Selected Poster prepared for presentation at the American Agricultural Economics Association  
Annual Meeting, Orlando, FL, July 27-29, 2008.*

*Copyright 2008 by Jae Bong Chang, Jayson L. Lusk, and F. Bailey Norwood. All rights reserved.  
Readers may make verbatim copies of this document for non-commercial purpose by any means,  
provided that this copyright notice appears on all such copies.*

**Abstract:** We compare the ability of three preference elicitation methods (hypothetical choices, non-hypothetical choices, and non-hypothetical rankings) and three discrete-choice econometric models (the multinomial logit, the independent availability logit, and the random parameter logit) to predict actual retail shopping behavior in three different product categories (ground beef, wheat flour, and dishwashing liquid). Overall, across all methods, we find a reasonably high level of external validity. Our results suggest that the non-hypothetical elicitation approaches, especially the non-hypothetical ranking, outperformed the hypothetical choice experiment in predicting retail sales. We also find that the random parameter logit can have superior predictive performance, but that the multinomial logit predicts equally well in some circumstances.

**Keywords:** contingent valuation, choice experiments, experimental economics, external validity, field experiment

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, and although a great deal has been learned, there are, with a few notable exceptions, very few studies examining the external validity of these methods. The skepticism surrounding stated and experimental willingness-to-pay values is perhaps justified, as practitioners advocating such methods have not adequately established the validity of the methods. This paper represents an attempt to empirically verify or refute such criticisms while simultaneously addressing some key issues for those individuals intimately working with revealed and stated preferences.

One of the key findings that has bolstered criticisms of valuation methods is that of hypothetical bias: research reveals that willingness-to-pay elicited from hypothetical, contingent valuation (CV) approaches almost always exceeds willingness-to-pay from non-hypothetical approaches (e.g., Cummings, Harrison, and Ruström, 1995; Fox et al. 1998; List and Gallet, 2001). 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, this need not be the case.

Non-hypothetical 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 (e.g., see Harrison and List, 2004; Levitt and List,

2007; List, 2006). For example, in an experimental setting, subjects 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, responses in a laboratory experiment may be influenced by the prices and available of substitutes outside the experiment.

By contrast, in a hypothetical experiment or survey, people may respond in a way that is more commiserate with how they would respond in a field setting where these artificial constraints are not present. In the context of valuation of public goods with CV methods, for example, Mitchell and Carson (1989) argue that external validity can be achieved with careful survey design and analysis. The counter-argument is that the lack of real monetary incentives in hypothetical experiments or surveys could very well lead to relatively poor external validity. People have little incentive to put cognitive effort into their decisions and face very little consequence from giving responses that deviate from true preferences.

Thus, neither hypothetical nor non-hypothetical valuation approaches are without their criticisms and it is far from clear which approach, if either, is reflective of people’s real world shopping behavior. However the intuitive appeal of non-hypothetical over hypothetical valuation may be, there is little empirical evidence to support such notions. Although a few studies have investigated the external validity of non-hypothetical preference elicitation methods (e.g., Brookshire, Coursey, and Schulze, 1987; Lusk, Pruitt, and Norwood, 2006; Shogren et al., 1999) or hypothetical contingent valuation methods (e.g., Vossler et al., 2003; Johnston, 2006), little is known about the relative performance of hypothetical and non-hypothetical approaches in predicting field shopping behavior. This issue is increasingly important as researchers are using real and hypothetical methods to estimate consumer preferences for food attributes for use in

agribusiness marketing and public policy decisions.

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 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 in the general literature on econometric predictions, more parsimonious models are often found to exhibit better 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 advocated by Louviere, Hensher, and Swait (2000) and used by Adamowicz et al., 1998 and Lusk, Roosen, and Fox (2003), b) non-hypothetical choice experiments of the type used by Alfnes et al. (2006), Ding, Grewal, and Liechty (2005), and Lusk and Shroeder (2004), and c) a

new non-hypothetical ranking experiment introduced by Lusk, Fields, and Prevatt (forthcoming). 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. Data collected from the non-hypothetical ranking method introduced by Lusk, Fields, and Prevatt (forthcoming) 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 various aspects related to the external validity of contingent valuation and experimental methods. These previous studies can be broadly grouped into four categories: 1) comparisons of hypothetical CV responses to real, field public referenda (e.g., Johnston, 2006; Vossler et al., 2003; Vossler and Kerkvliet, 2003), 2) comparisons of hypothetical CV 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) comparisons of non-hypothetical experimental behavior to real, field shopping behavior (e.g., Brookshire, Coursey, and Schulze, 1987; Lusk, Pruitt, and Norwood, 2006; Shogren et al., 1999), and 4) investigations into hypothetical bias involving comparisons of non-hypothetical experimental behavior to either hypothetical experimental behavior or hypothetical CV responses (e.g., Cummings, Harrison, and Rutström, 1995; Fox et al., 1998; List and Shogren, 1998).

Despite these studies and the fact that stated preferences are increasingly used to explain consumers' behavior, there remain important questions related to external validity. Much of the literature comparing hypothetical CV responses to real, field public referenda and to indirect valuation methods suggest a reasonably high level of convergent validity – i.e., the CV responses seem to map reasonably well to observed behavior depending on how one chooses to model and code “indifferent” and “don't know” responses with the CV method (e.g., see Vossler and Kerkvliet, 2003 and Carson et al., 1996). These findings, however, stand in stark contrast with the standard findings from the papers on hypothetical bias that indicate large differences between CV-type responses and behavior when money is on the line (see List and Gallet, 2001 for a review). This begs the question as to *which* approach, real or hypothetical, exhibits the highest level of external validity. Further, whereas much of the focus of CV studies has been on the external validity of the method at measuring the value of public goods via referenda-type questions, there is now a great deal of interest in using choice-experiment methods and in 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 (e.g. see Carson and Groves, 2007).

The only study of which we are aware that has directly tackled the question of the relative predictive performance of hypothetical vs. real responses was that by Shogren et al. (1999), who compared survey responses from a mail survey and behavior in a non-hypothetical experimental 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.

Shogren et al. (1999) found that 80% of 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, when the irradiated chicken breast was sold at a premium, the survey and experimental results predicted market share in a subsequent retail trial quite well. 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. Consistent with the literature on hypothetical bias, Shogren et al. (1999) found that the implied WTP for irradiated chicken was higher for the hypothetical survey data than the experimental data, however the difference was not quantitatively large as there was only about a \$0.06 difference between real and hypothetical implied valuations.

Results from Shogren et al. (1999) suggest mixed results regarding the external validity of both hypothetical and non-hypothetical responses and 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 important ways. First, we utilize what is becoming a standard valuation approach: the choice experiment (Louviere, Hensher, and Swait, 2000). Second, like Shogren et al. (1999), we compare real and hypothetical responses in a choice experiment to actual retail sales, but unlike Shogren et al. (1999), we carry out formal statistical tests to determine which method is superior. Third, unlike most previous studies, rather than focusing on a single good, we compare predicted and actual market share for fifteen 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 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*

A random sample of people from Stillwater, OK 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 research study in exchange for \$40. Thus, this application utilized a representative sample of respondents such that participants in the laboratory experiments were reflective of typical shoppers in the local grocery store. This does not mean that no students participated, only that they participated proportionately to their share of the local population. Once a contacted

individual agreed to participate, they were mailed a reminder note and a map to the study's location. Upon arrival at the study site, subjects were randomly divided into one of three experimental treatments: hypothetical choice, non-hypothetical choice, or non-hypothetical ranking. In total, forty-seven consumers participated in the hypothetical choice treatment, forty-six people participated in the non-hypothetical choice treatment, and another forty-two subjects were assigned to the non-hypothetical ranking treatment.

Regardless of the treatment to which an individual was assigned, they were requested to investigate 15 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 15 products were grouped into three product categories: dishwashing liquid, ground beef, and wheat flour. In each product category were three pre-existing brands and one new brand that was not available for sale in the local market. The three new products were: 1) Eco-Plus, an "environmentally friendly" dishwashing liquid, 2) Cattle Tracks, an organic ground beef brand, and 3) GO Organic wheat flour, an organic and regionally grown brand.

The products used in the experiment were chosen to mirror that offered in a local grocery that agreed to participate in the research. The local store sold each of these 15 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 closely followed the approach 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 completely uncorrelated with each other across the design. Figure 1 shows an example of a discrete choice question for each of the three product categories. The order with 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.

The third treatment utilized the non-hypothetical conjoint ranking approach introduced by Lusk, Fields, and Prevatt (forthcoming). 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. Examples of non-hypothetical ranking valuation questions are presented in figure 2.

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 the numbers 1-15 were drawn, the participant purchased the product they ranked first, if the numbers 16-25 were drawn, they purchased the product they ranked second, if the numbers 26-30 were drawn, they purchased the product ranked third, if the numbers 31-34 were drawn, they purchased the product ranked fourth, and if 35 or 36 were drawn, they purchased the product ranked last. In this way, there was a higher chance a participant purchased a product they gave a lower rank. 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, OK 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 sell 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

$$(1) \quad U_{ij} = V_{ij} + e_{ij}$$

where  $V_{ij}$  is a deterministic component and  $e_{ij}$  is an iid stochastic component. In this application,

the deterministic portion of the utility function can be expressed as

$$(2) \quad V_{ij} = a_{ij} + a_{price} P_{ij}$$

where  $a_j$  are alternative specific constants indicating utility for alternative  $j$  relative to an omitted option,  $a_{price}$  represent the marginal utility of price, and  $P_{ij}$  is the price of alternative  $j$  for consumer  $i$ .

### *Multinomial Logit Model*

Assuming the  $e_{ij}$  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:

$$(3) \quad \text{Prob}\{j \text{ is chosen}\} = \frac{\exp(V_{ij})}{\sum_{k=1}^J \exp(V_{ik})}.$$

Equation (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:

$$(4) \quad \prod_{j=1}^{J-1} \frac{e^{V_{ij}}}{\sum_{k=j}^J e^{V_{ik}}}$$

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. This assumption, however, might be overly restrictive. Some people may only consider a subset of all available options. If so, the MNL formula in (3) will be incorrect for such an individual as the choice probabilities are calculated by summing over the utility of all  $J$  goods. That some people never consider a particular choice option is equivalent to saying they place a value of negative infinity on that option. Although some people may never consider a particular alternative, others may have some likelihood of choosing it. The MNL, which assumes that every consumer considers every alternative, cannot accommodate such behavior.

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 of 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). The estimator consists of two stages. The first stage consists of forming a consideration set and the second stage consists of choosing an alternative from the given consideration set. Given this assumption, Louviere, Hensher, and Swait (2000) show that the probability of individual  $i$  choosing option  $j$  in the IAL model is



$$(5) \quad \text{Prob}\{j \text{ is chosen}\} = \sum_{C_i \subseteq C} \text{Prob}(j|C_i) \times \text{Prob}(C_i) = \frac{\exp(V_{ij})}{\sum_{k \in C_i} \exp(V_{ik})} \times \frac{\prod_{j \in C_i} A_{ij} \prod_{j \in C - C_i} (1 - A_{ij})}{1 - \prod_{j \in C} (1 - A_{ij})}$$

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 (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

$$(6) \quad A_{ij} = \frac{1}{(1 + \exp(-b_{ij}))}$$

where  $b_{ij}$  are alternative-specific constants. 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. Then, for the product ranked second, it was assumed it would have been chosen as most preferred out of the remaining four options. Then, 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.<sup>1</sup>

---

<sup>1</sup> Although some might object to this re-coding, it is important to note that estimating the rank ordered logit in equation (4) yields the same result as estimating the MNL in equation (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.

### *Random Parameter Logit Model*

The MNL assumes preference homogeneity in the sample. This implies that all coefficients of the utility expression in (1) are 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 (usually normal distributions).

The RPL is implemented by specifying the alternative specific constants shown in equation (2) as

$$(7) \quad a_{ij} = \bar{a}_j + s_j n_{ij}$$

where  $\bar{a}_j$  is the population mean alternative specific constant for option  $j$ ,  $s_j$  is the standard deviation of the distribution of the coefficient  $a_{ij}$  around the population mean, and  $n_{ij}$  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

$$(8) \quad \text{Prob}\{j \text{ is chosen}\} = \int \frac{\exp(V_{ij})}{\sum_{k \in C_i} \exp(V_{ik})} f(a_i) da_i$$

where  $f(a_i)$  is the density of coefficients  $a_i$ . Because equation (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 (3), (5), and (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:

$$(9) \quad OSLLF = \sum_{j=1}^J TM_j \ln(EM_j),$$

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 is slightly more involved. 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  $\bar{a}_j$  and  $s_j$  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: 1,000 simulated individuals for each of 1,000 mean parameter draws), 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 25<sup>th</sup> and 975<sup>th</sup> highest mean MSE/OSLLF values across the 1,000 mean parameter draws. 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 1 reports estimates of nine MNL models (three elicitation methods \* three product categories). For each elicitation methods 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 all positive except for the King Arthur constant in the non-hypothetical choice treatment, 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 ( $p\text{-value} < 0.01$ ) for all nine models shown in table 1.<sup>2</sup>

Table 2 represents 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 alternatives 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 emerge 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  $\beta_4 < 0$ ).

Table 2 also reports results for the means 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 dishwashing liquid and wheat flour in non-hypothetical choice method, implying a significant amount of preference heterogeneity. That the magnitude of the standard deviation of preferences for King Arthur flour ( $\alpha_2$ ) in non-hypothetical method is extremely large, is indicative of the fact that only one subject chose this option in this particular

---

<sup>2</sup> Caution should be taken in directly comparing coefficients across elicitation methods as it involves a comparison of both the utility parameters and the error variance (Swait and Louviere 1993).

treatment.

Table 3 reports the predicted market shares for each product by experimental treatment and econometric model. The last column in table 3 reports the market share from actual store sales. 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 for products from the MNL, non-hypothetical ranking treatment for dishwashing liquid never diverge from the true market share by more than 3% for any product.

Table 4 contains the key comparisons between the methods. Shown in table 4 are the MSE and OSLLF for each experimental treatment, estimation method, and product category. To compare predicted market shares between methods, two criteria were used to rank methods: MSE and OSLLF. Focusing first on the MSE criteria, for which a lower value is preferred, we find that for the dishwashing liquid category in the MNL model, MSE for the non-hypothetical choice method is lowest at 0.001. However, for ground beef and whole wheat flour, MSE for the non-hypothetical ranking treatment is lowest at 0.07 and 0.105, respectively. For the IAL model, 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, 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 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 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 value across the MNL and RPL models is 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 value across the MNL and RPL models is 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 attempt to further summarize the findings in two ways. First, the last three rows in table 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 the RPL yields lower MSE and higher OSLLF at the  $p < 0.01$  level for the non-hypothetical ranking method when data is 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 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 4 to test the hypothesis the three key variables (product category, elicitation method, and econometric model) and their interactions affect MSE and OSLLF. Results of the ANOVA are shown in table 5. Results 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 5 confirms that 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 predictive performance varied across three elicitation methods (hypothetical choice, non-hypothetical choice, and non-hypothetical ranking), across three econometric models (the MNL, IAL, and RPL) and three product categories (dishwashing liquid, ground beef, and whole wheat flour). Our econometric results 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 also 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 out-of-sample predictions as well. However, this is

not always be 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 restricted (e.g., see Haab and Hicks, 1997; Swait and Ben-Akiva. 1987).

A final note to those wary of survey and experimental methods is in order. As shown in table 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 resulting 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. 0.40 for Joy, 0.13 vs. 0.19 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.

## 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.
- 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.
- 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.
- 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.
- 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.
- Cummings, R.G., G.W. Harrison, and E.E. Rutström. 1995. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review* 85:260-66.
- Ding, M.R. Grewal, and J. Liechty. 2005. "Incentive-Aligned Conjoint Analysis." *Journal of Marketing Research* 42:67-83.
- 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.
- 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.

- 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 Tell Us About the Real World?" *Journal of Economic Perspectives* 71:153-74.
- 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.
- List, J.A., and J. Shogren. 1998. "Calibration of the Difference between Actual and Hypothetical Valuations in a Field Experiment." *Journal of Economic Behavior and Organization* 37:193-205.
- 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: Cambridge University Press.

- Lusk, J.L., D. Fields, and W. Prevatt. Forthcoming. "An Incentive Compatible Conjoint Ranking Mechanism." *American Journal of Agricultural Economics*, in press.
- 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.
- Mitchell, R. C., and R. T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Baltimore: John Hopkins University Press.
- Murphy, E., B. Norwood, and M. Wohlgenant. 2004. "Do Economic Restrictions Improve Forecasts?" *Journal of Agricultural and Applied Economics* 36:549-58.
- 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. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.
- 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.

**Shopping Scenario 1: Please check the ONE item you prefer**

	Palmolive: Original Scent (25 fl oz) <b>\$3.00</b> ↓	Dawn: Original Scent (25 fl oz) <b>\$2.00</b> ↓	Joy: Original Scent (25 fl oz) <b>\$4.00</b> ↓	Eco Plus (28 fl oz) <b>\$4.00</b> ↓	<b>NONE</b>  <b>\$0.00</b> ↓
<b>I Choose (check one)</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Shopping Scenario 6: Please check the ONE item you prefer**

	Diet Lean Ground Beef <b>\$2.00</b> ↓	Fresh Ground Beef <b>\$4.00</b> ↓	Lean Ground Beef <b>\$3.00</b> ↓	Cattle Tracks Organic Ground Beef <b>\$4.00</b> ↓	<b>NONE</b>  <b>\$0.00</b> ↓
<b>I Choose (check one)</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Shopping Scenario 11: Please check the ONE item you prefer**

	Gold Medal Whole Wheat Flour <b>\$3.00</b> ↓	Go Organic Whole Wheat Flour <b>\$4.00</b> ↓	King Arthur Whole Wheat Flour <b>\$2.00</b> ↓	Hodgson Mill Whole Wheat Flour <b>\$2.00</b> ↓	<b>NONE</b>  <b>\$0.00</b> ↓
<b>I Choose (check one)</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

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



**Scenario 1: Please rank the following goods from the most preferred to least preferred (1=most preferred, 5=least preferred)**

	Palmolive: Original Scent (25 fl oz) <b>\$3.00</b> ↓	Dawn: Original Scent (25 fl oz) <b>\$3.00</b> ↓	Joy: Original Scent (25 fl oz) <b>\$3.00</b> ↓	ECO Plus (28 fl oz) <b>\$3.00</b> ↓	<b>NONE</b> <b>\$0.00</b> ↓
<b>Ranking</b>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Scenario 6: Please rank the following goods from the most preferred to least preferred (1=most preferred, 5=least preferred)**

	Diet Lean Ground Beef <b>\$4.00</b> ↓	Fresh Ground Beef <b>\$4.00</b> ↓	Lean Ground Beef <b>\$4.00</b> ↓	Cattle Tracks Organic Ground Beef <b>\$3.00</b> ↓	<b>NONE</b> <b>\$0.00</b> ↓
<b>Ranking</b>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Scenario 11: Please rank the following goods from the most preferred to least preferred (1=most preferred, 5=least preferred)**

	Gold Medal Whole Wheat Flour <b>\$3.00</b> ↓	Go Organic Whole Wheat Flour <b>\$4.00</b> ↓	King Arthur Whole Wheat Flour <b>\$2.00</b> ↓	Hodgson Mill Whole Wheat Flour <b>\$2.00</b> ↓	<b>NONE</b> <b>\$0.00</b> ↓
<b>Ranking</b>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

**Figure 2.** Examples of questions for the non-hypothetical ranking method

**Table 1.** Multinomial Logit Model Estimates by Product Category and Elicitation

Method

Commodities	Elicitation Method		
	Hypothetical Choice	Non-hypothetical Choice	Non-hypothetical Ranking
<i>Dishwashing Liquid</i>			
$a_1$ , Dawn	4.81* <sup>a</sup> (0.48) <sup>b</sup>	1.63* (0.82)	3.09* (0.13)
$a_2$ , Joy	4.06* (0.47)	1.38 (0.70)	2.71* (0.12)
$a_3$ , Palmolive	3.76* (0.45)	1.79* (0.78)	2.74* (0.13)
$a_4$ , Eco-Plus	4.16* (0.45)	1.66 (0.85)	2.53* (0.12)
$a_{price}$	-1.45* (0.15)	-1.66* (0.33)	-1.13* (0.04)
Log likelihood	-294.29	-164.97	-884.43
<i>Ground Beef</i>			
$a_1$ , Fresh	4.02* (0.44)	1.15 (0.85)	1.94* (0.23)
$a_2$ , Lean	5.15* (0.45)	2.76* (0.65)	2.43* (0.24)
$a_3$ , Diet Lean	3.99* (0.44)	2.79* (0.75)	2.02* (0.24)
$a_4$ , Organic	3.64* (0.44)	2.47* (0.77)	1.77* (0.23)
$a_{price}$	-1.28* (0.14)	-1.85* (0.30)	-0.82* (0.07)
Log likelihood	-285.32	-184.16	-900.99
<i>Whole Wheat Flour</i>			
$a_1$ , Hodgson Mill	4.56* (0.50)	2.69* (0.80)	2.28* (0.24)
$a_2$ , King Arthur	3.61* (0.49)	-0.41 (1.29)	2.46* (0.24)
$a_3$ , Gold Medal	5.78* (0.53)	2.69* (0.78)	2.57* (0.25)
$a_4$ , GO Organic	4.70* (0.51)	1.58 (0.90)	2.45* (0.24)
$a_{price}$	-1.80* (0.16)	-1.93* (0.34)	-0.97* (0.07)
Log likelihood	-261.17	-140.38	-882.66
Number of Observations	235	230	215

<sup>a</sup> One asterisk (\*) denotes values that are statistically significant at the 0.05 level.<sup>b</sup> Numbers in parentheses are standard errors.

**Table 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
<i>Utility function</i>																		
$a_1$	-1224* (0.72) <sup>a</sup>	156 (0.93)	2.65* (0.74)	1332* (2.49)	-560 (1755)	8.52* (0.00)	0.30 (0.93)	1.31 (0.95)	1.25 (0.77)	2942* (10.12)	631 (885)	582* (0.00)	-1105* (1.18)	8.22* (1.66)	5.98* (1.23)	17.11* (4.53)	21.25 (12.94)	4.51* (0.00)
$a_2$	-1213* (0.74)	2.04* (0.84)	2.04* (0.87)	11.04* (2.08)	-25.46 (29.02)	5.42* (0.85)	3.60* (1.14)	3.81* (0.82)	3.52* (0.80)	40.29* (13.95)	14.31* (6.62)	7.09* (0.00)	-11.86* (1.12)	1.20 (1.58)	1.47* (0.67)	12.35* (4.36)	-1.692 (5.0+8)	2.85* (0.00)
$a_3$	-1268* (0.74)	2.57* (0.97)	2.89* (0.86)	10.38* (1.96)	13.88 (10.37)	4.94* (0.00)	0.84 (1.00)	3.80* (0.95)	3.01* (0.76)	27.05* (9.49)	12.96* (5.68)	4.90* (0.00)	-7.80* (1.51)	4.84* (1.27)	2.40* (0.66)	21.86* (5.61)	-1.26 (7.11)	2.95* (0.00)
$a_4$	-1293* (0.67)	5.50* (0.81)	8.35* (2.44)	11.80* (2.44)	18.09 (14.73)	3.35* (0.00)	-0.13 (0.89)	6.37* (0.80)	7.44* (0.76)	1.75 (4.76)	20.30* (8.27)	4.52* (0.01)	-11.02* (1.15)	2.62* (1.28)	1.26* (0.59)	17.39* (4.48)	8.12 (11.22)	7.53* (0.00)
$a_{price}$	-1.64* (0.22)	-1.62* (0.37)	-1.52* (0.31)	-4.37* (0.77)	-15.84 (11.30)	-2.41* (0.26)	-1.85* (0.28)	-1.92* (0.35)	-1.43* (0.27)	-10.26* (3.37)	-12.12* (4.25)	-2.19* (0.00)	-2.91* (0.54)	-2.42* (0.55)	-1.08* (0.22)	-6.86* (1.69)	-14.67 (8.61)	-2.04* (0.00)
<i>Availability/standard deviation parameters<sup>c</sup></i>																		
$b_1/\sigma_1$	17.37* (0.11)	22.87* (0.07)	19.61* (0.08)	5.07* (1.08)	27.48 (24.15)	4.43* (0.19)	16.38* (0.09)	22.30* (0.07)	18.76* (0.08)	16.39* (6.43)	11.96* (5.78)	4.56* (0.00)	16.42* (0.17)	-2.11* (0.25)	-2.53* (0.29)	8.13* (2.64)	13.97 (7.84)	4.79* (0.00)
$b_2/\sigma_2$	17.49* (0.08)	22.39* (0.07)	19.58* (0.08)	4.55* (1.16)	41.42 (34.05)	3.71* (0.58)	16.85* (0.08)	20.24* (0.07)	18.69* (0.08)	11.85* (4.43)	19.60* (6.79)	6.13* (0.00)	16.43* (0.08)	19.30* (0.07)	18.53* (0.07)	8.50* (2.42)	912.43 (2.0+8)	8.08* (0.00)
$b_3/\sigma_3$	-17.14* (0.08)	-19.65* (0.07)	-18.67* (0.08)	4.71* (1.62)	27.64 (21.10)	7.16* (0.00)	-16.28* (0.09)	-19.78* (0.07)	-17.94* (0.08)	14.92* (5.81)	13.95* (4.97)	6.12* (0.00)	-16.39* (0.10)	-18.55* (0.07)	-17.89* (0.07)	7.08* (1.87)	30.17 (19.96)	4.74* (0.00)
$b_4/\sigma_4$	16.90* (0.18)	-2.55* (0.31)	-1.91* (0.23)	6.21* (1.46)	20.07 (13.95)	3.01* (0.00)	16.02* (0.18)	-2.25* (0.27)	-1.82* (0.22)	31.90* (13.04)	6.50* (2.41)	4.91* (0.02)	16.58* (0.18)	21.04* (0.07)	18.53* (0.07)	7.41* (2.45)	15.89 (11.78)	2.96* (0.00)
$b_{None}$	-1.61* (0.18)	17.28* (0.22)	17.06* (0.16)				-1.91* (0.23)	17.63* (0.44)	17.23* (0.20)				-1.31* (0.16)	16.85* (0.11)	17.07* (0.12)			
LL <sup>d</sup>	-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

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

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

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

<sup>a</sup> Numbers in parentheses are standard errors determined via parametric bootstrapping.

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

	MNL		IAL		RPL	
	MSE <sup>a</sup>	OSLLF <sup>b</sup>	MSE	OSLLF	MSE	OSLLF
<i>Dishwashing Liquid</i>						
Hypothetical	0.028 [0.006,0.078] <sup>c</sup>	-1.065 [-1.142,-1.028]	0.043 [0.027,0.133]	-3.336 [-3.488,-3.238]	0.014 [0.003,0.129]	-1.054 [-1.253,-1.016]
Nonhypothetical	0.001 [0.001,0.090]	-1.006 [-1.141,-1.002]	0.088 [0.026,0.277]	-3.597 [-3.863,-3.483]	0.124 [0.013,0.444]	-1.298 [-2.615,-1.035]
Real ranking	0.002 [0.001,0.010]	-1.019 [-1.035,-1.012]	0.041 [0.026,0.116]	-3.517 [-3.641,-3.386]	0.032 [0.032,0.032]	-1.052 [-1.052,-1.052]
<i>Ground Beef</i>						
Hypothetical	0.053 [0.027,0.098]	-1.208 [-1.293,-1.151]	0.407 [0.267,0.482]	-6.108 [-6.721,-5.580]	0.040 [0.030,0.106]	-1.196 [-1.321,-1.173]
Nonhypothetical	0.027 [0.002,0.118]	-1.139 [-1.343,-1.097]	0.281 [0.113,0.384]	-6.216 [-6.990,-6.108]	0.004 [0.003,0.172]	-1.126 [-2.351,-1.098]
Real ranking	0.007 [0.004,0.016]	-1.143 [-1.164,-1.130]	0.239 [0.133,0.335]	-5.876 [-6.210,-5.623]	0.005 [0.005,0.005]	-1.140 [-1.140,-1.139]
<i>Whole Wheat Flour</i>						
Hypothetical	0.441 [0.343,0.524]	-2.521 [-2.877,-2.212]	0.700 [0.425,0.809]	-6.257 [-6.949,-5.709]	0.251 [0.046,0.655]	-1.873 [-7.422,-1.428]
Nonhypothetical	0.251 [0.228,0.398]	-2.413 [-3.330,-2.244]	0.221 [0.163,0.375]	-6.084 [-6.336,-6.067]	0.191 [0.047,0.425]	-1.871 [-1.122,-1.676]
Real ranking	0.105 [0.074,0.143]	-1.561 [-1.641,-1.491]	0.229 [0.185,0.369]	-6.053 [-6.184,-5.973]	0.010 [0.010,0.010]	-1.364 [-1.364,-1.364]
<i>Total</i>						
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	<b>0.114</b>	<b>-3.722</b>	<b>0.509</b>	<b>-15.447</b>	<b>0.047</b>	<b>-3.556</b>

<sup>a</sup> MSE is mean squared error between predicted and actual market share summed across each product

<sup>b</sup> OSLLF is the estimated likelihood function value observed at actual market share values.

<sup>c</sup> Numbers in brackets are 95% confidence intervals determined via parametric bootstrapping.

**Table 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 <sup>a</sup>	OSLLF <sup>b</sup>
Product Category	2	33.11 <sup>*c</sup>	162.31 <sup>*</sup>
Elicitation Method	2	13.52 <sup>*</sup>	4.79 <sup>*</sup>
Econometric Model	2	22.52 <sup>*</sup>	1481.89 <sup>*</sup>
Product Category * Elicitation Method	4	7.81 <sup>*</sup>	2.6
Product Category * Econometric Model	4	6.28 <sup>*</sup>	58.28 <sup>*</sup>
Elicitation Method * Econometric Model	4	2.01	0.5
R <sup>2</sup>		0.96	0.99
Number of Observations		27	27

<sup>a</sup>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.

<sup>b</sup>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.

<sup>c</sup>One asterisk indicates the null hypothesis of no effect of the independent variable can be rejected at 0.05 level or lower.