Preference Inconsistencies of a Rational Decision Maker

Bachir Kassas, Marco A. Palma and Yvette Zhang
Department of Agricultural Economics, Texas A&M University, College Station, TX 77840, USA

Selected paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, February 6-9,2016, San Antonio, Texas

Preference Inconsistencies of a Rational Decision Maker

Bachir Kassas[a,*], **Marco A. Palma**[a] **and Yvette Zhang**[a]

[a] *Department of Agricultural Economics, Texas A&M University, College Station, TX 77840, USA*

Abstract:

The longstanding dispute over the accuracy of stated preference methods in eliciting the true valuations of individuals has stimulated interest in analyzing preference inconsistencies between revealed and stated preference mechanisms. This paper uses *preference orderings* to provide a more robust comparison between revealed and stated preferences and assess the validity of the latter. This is done by comparing an incentive compatible auction experiment (recoded as implied ranks) with a ranking procedure. Partial ranking models are constructed to examine consumer preferences under the two valuation mechanisms for the most preferred and the least preferred alternatives in order to provide a more detailed analysis. The stability and symmetry of parameters was tested and systematic differences between the models were analyzed in order to measure the extent of preference inconsistencies between the auction exercise and ranking procedure. Furthermore, the predictive power of the models was calculated to evaluate the relative reliability of each mechanism. The results provide robust evidence that individuals often employ different behavioral rules under the two elicitation mechanisms, especially when expressing mild feelings about certain alternatives. Compared to the more accurate auctions mechanism, the ranking exercise seems to perform fairly well only when eliciting preferences over the best ranked alternative.

JEL Classification: D12

Key Words: auctions, choice-ranking**,** ordinal data, parameter stability, parameter symmetry, preference inconsistency, revealed preferences, stated preference

1. Introduction

Consumer preferences have been of prominent interest to researchers in several fields. In fact, a significant amount of resources has been directed towards the construction and analysis of different preference elicitation mechanisms. Furthermore, different theories have been developed to improve our understanding of the consumer's decision process and the factors that influence his valuation for different goods and services.

There are two general approaches to elicit consumer preferences: 1) stated preference mechanism; and 2) revealed preference mechanism. Discrete choice-ranking experiments are among the most commonly used stated preference mechanisms. Choice experiments have been applied to elicit preferences for various products and services including household appliances, clean fuel vehicles, travel choices, alternative therapies, and environmental assets (Revelt and Train 1998, Bunch et al. 1993, Hensher 1994, McNeil et al. 1982, Hanley et al. 1998). Moreover, several partial rankings have been used to elicit consumer preferences including top ranks, bottom ranks and best-worst ranking (Bockenholt 1992, Pavan and Todeschini 2004, Hensher and Ho 2015).

Contrary to the stated preference approach, which relies on surveys and hypothetical methods, revealed preference mechanisms employ incentive compatible techniques to elicit consumer preferences in non-hypothetical markets. Auctions are the predominantly used form of revealed preference methods. Different types of auctions have been utilized in laboratory experiments to reveal consumer valuations including Dutch auctions, English auctions, Vickrey first-, second-, and n^{th} price auctions, and Becker-DeGroot-Marschak (BDM) auctions (Katok and Roth 2004, Rutstrom 1998, Kagel and Levin 1993, Shogren et al. 2001, Lusk 2003).

There is a longstanding dispute over the accuracy of stated preference methods in eliciting individuals' true underlying preferences. Some proponents of stated preferences argue in favor of the validity of the approach (Wardman 1988). However, stated preference methods have faced major criticism and are treated with skepticism by many researchers. Murphy et al. (2005) reported a series of twenty eight studies which compared results of stated preference elicitation mechanisms with actual values. They concluded that stated preference mechanisms were biased in the majority of those studies as subjects tend to overstate their true values by a factor of two or three. Furthermore, List and Gallet (2001) conducted a meta-analysis over twenty nine experimental studies that dealt with willingness-to-pay (WTP) and willingness-to-accept (WTA) estimates. They report similar results in that subjects tend to overstate their values in stated preference methods by a factor of 3.

In this article, we provide a robust comparison between revealed and stated preferences which will help generate a more reliable assessment of the accuracy of the latter. This is done by comparing consumer preference *"orderings"* between a ranking procedure and an incentive compatible auction mechanism for seven new pomegranate products. Specifically, we test for the existence of preference inconsistencies between the two valuation methods. The main objectives of this article are: 1) to identify and validate *preference ordering* inconsistencies between rankings and experimental auctions; 2) to empirically assess the results of full and partial ranking across the two elicitation methods; 3) to check for systematic differences in preference orderings between the auction exercise and ranking procedure; 4) to compare the predictive efficiency of each mechanism; and 5) to test the stability and symmetry of parameters across the two valuation methods in order to determine whether individuals follow the same decision rules under each mechanism.

While most comparisons have focused on the actual consumer valuation like willingness to pay, our study uses consumer *preference orderings* to explore this issue (i.e. is the most preferred option the same across the elicitation mechanisms? what about the second best option, or the worst?). Valuations are a cardinal measure associated with a cardinal utility while ranking orders are ordinal in nature and generate an ordinal utility. Thus, valuations are susceptible to bias and noise since they are a more exact measure and might not reflect the individual's true realized utility as they may be influenced by exogenous factors. Therefore, in order to provide a better insight over the individual's self-inconsistencies we use ordinal data by recoding the auction bids into implied ranks. We find robust evidence that individuals often employ different behavioral rules under the two elicitation mechanisms, especially when expressing mild feelings about certain alternatives. Compared to the more accurate auction mechanism, arriving from its incentive compatibility, the ranking exercise seems to perform fairly well only when eliciting preferences over the most preferred alternative. However, the accuracy of the ranking exercise in eliciting true valuations substantially decreases when other partial rankings are included. This is evident in the drastic decrease in the predictive power of the models based on the ranking exercise compared to those of the auction mechanism and the fact that subjects' responses differ systematically under both mechanisms when partial rankings are considered.

The rest of the paper is organized as follows: Section two briefly reviews the literature of preference inconsistencies. Section three describes the experimental procedure and the process of generating the data. Section four presents the analysis framework and econometric models followed by a discussion of the results in section five. The last section highlights the important findings of this study and concludes.

2. Literature review on preference inconsistencies

Preference inconsistencies have been studied extensively in the literature. Grether and Plott (1979), and Tversky et al. (1990) focused on finding an explanation for this phenomenon. Furthermore, Mowen and Gentry (1980), and Mellers et al. (1992) worked on constructing experimental designs to deal with those inconsistencies, where they presented the task under different contexts and tried skewed distributions of expected value. However, the attempts to overcome this problem have not been successful in providing a satisfactory solution (Lichtenstein and Slovic 1973, Grether and Plott 1979, Berg et al. 1985, Chu and Chu 1990).

The inability to estimate consistent preferences across different valuation mechanisms has driven interest towards analyzing the underlying forces that cause these inconsistencies. Researchers started comparing consumer preferences under different valuation mechanisms such as choice versus ranking (Caparros et al. 2008), rating versus ranking (Harzing et al. 2009, Alwin and Krosnick 1985, Sayadi et al. 2005, and Mackenzie 1993) and auction/bids versus choice (Su et al. 2011, Lusk and Schroeder 2006, Corrigan et al. 2009). In addition, Boyle et al. (2001) compared rating, choice, ranking, and reordered ranking, while Siikamaki and Layton (2007) used contingent valuation (CV) versus contingent rating/ranking (CR).

The majority of comparisons focused mainly on consumer valuations and willingness to pay (WTP) estimates. For instance, Su et al. (2011) reported that WTP in auctions is higher than that in choice experiments. They also found that WTP in CR is higher than that in CV. Conversely, Siikamaki and Layton (2007) asserted that CV generates WTP estimates that are one order of magnitude bigger than those under CR. Moreover, Corrigan et al. (2009) concluded that WTP estimates calculated under open-ended choice experiments exhibit less affiliation across rounds than the estimates obtained under uniform price auctions making the latter a more favorable preference elicitation mechanism.

Researchers have exploited several different comparison methods in their analyses, the most ubiquitous of which was the comparison of parameter estimates. When comparing choice experiments to ranking exercises, Boyle et al. (2001) found persistent differences in parameter estimates. In contrast, Morrison and Boyle (2001), and Capparos et al. (2008) analyzed the same elicitation mechanisms and their results suggested that choice experiments and ranking exercises, when recoded as choice, provide statistically similar parameter vectors.

The close inspection and analysis of preference inconsistencies has enabled researchers to devise various explanations for the phenomenon. Loomes and Sugden (1983) attributed the inconsistencies in preference ordering between auction/bids and choice experiments to an underlying regret-utility hypothesis. They argued that choice experiments, which require the rejection of all non-chosen alternatives, impose more cognitive dissonance on the participant. Given proper assumptions about the underlying utility, this might lead participants to choose an alternative with a lower bid.

Some other plausible causes of preference inconsistencies include experimental design and information treatments. Capparos et al. (2008) argue that it is hard to discern if the inconsistencies are actually caused by the experimental design itself. In addition, it is believed that different designs stimulate different attributes of the alternatives and that confusion and misperception lead to inconsistencies especially when the participants are not well trained in advance (Lusk and Schroeder 2006, Plott and Zeiler 2005). On the other hand, respondents usually place a higher weight on information treatments in the experiment especially when providing a preference ranking. Hence, more information about the alternatives helps generate more consistent preferences (Su et al. 2011, McAdams et al. 2013).

The analysis of preference reversals, a special case of inconsistency, has helped provide more insightful explanations of this behavior. According to the New Palgrave Dictionary of Economics (2008), preference reversal is a wide spread behavioral property. It refers to situations where subjects report opposite or contradictory preference for two alternatives when facing different elicitation methods or contexts. (See Kim et al. (2012) for an extensive review of preference reversals).

A predominant cause of preference reversals is the change in consumers' views at different stages of the decision process. For example, Tversky et al. (1988) argued that participants weighed the attributes lexicographically, which caused them to switch their preference between stages. Alternatively, Mellers et al. (1992) asserted that changes in the way attributes are combined, for example additively or multiplicatively, resulted in significant discrepancies which gave rise to inconsistencies. Furthermore, Goldstein and Einhorn (1987) attributed the preference reversals to a change in how a formed valuation is expressed or translated into a response.

Several other causes of preference reversals were provided by Grether and Plott (1979), who categorized them as economic or psychological. The main economic explanations are miss-specified incentives, income effect and resolution of indifferences. Miss-specification of incentives suggests that the choice behavior depends on the level of interest of the respondent. On the other hand, income effect is exhibited when preferences depend on the level of income, while resolution of indifferences refers to situations where the participant develops a systematic way to deal with indifferences.

From a psychological perspective, the leading explanations of preference inconsistencies are: 1) strategic response, where the subject has a true price and strategically bids higher or lower than this price depending on the situation; 2) information processing – decision cost, in which participants anchor on the main attribute in an attempt to shed some of the cost associated with evaluating the alternatives; and 3) information processing – response mode and easy justification, where certain phrases and response modes impact the way subjects interpret the information.

3. Experimental design and data

A total of 202 individuals were selected to participate in a non-hypothetical experiment. Care was taken to ensure that participants were representative of grocery shoppers in the area (nonstudents). Local newspaper and internet ads were used to recruit the subjects to the experiment. The participants were placed in one of eight sessions where each session included 20-30 subjects. The subjects were randomized across the sessions in a way that mimicked the demographic and socioeconomic characteristics of U.S grocery shoppers (Carpenter and Moore 2006). Each participant received a $35 participation fee, in cash, at the end of the experiment excluding the amount of any purchases made during the experiment. Table 1 shows the demographic and socioeconomic summary statistics of the participants.

The experiment consisted of two parts. In the first part, subjects engaged in a non-hypothetical ranking procedure where they were presented with the different alternatives and were asked to rank them in order of preference. A rank of one was assigned to the most preferred alternative meaning that the highest ranked alternative received the lowest numerical value. The second part of the experiment involved an 11th price sealed bid Vickrey (1961) non-hypothetical, incentive compatible auction during which the market price of the products was determined. In this form of auction mechanism, the 11^{th} highest price is taken as the market price and the 10 highest bidders, those who bid higher than the market price, would buy the product at the market price. Since there were around 20 participants in each session, the market price was relatively close to the median bid.

The non-hypothetical ranking procedure was performed in two stages. In the first stage, each participant was asked to select his consideration set (the set of alternatives he was actually interested in or would consider purchasing). This meant that participants could have a different number of alternatives in their consideration sets depending on their preferences. In the second stage, the participants would rank-order the alternatives in their respective consideration sets. The two stages were performed simultaneously by each subject. Equivalently, in the auction exercise, the subjects communicated their interest in alternatives by bidding a positive amount for them. This meant that only those alternatives that received positive bids were included in the consideration set in the auction exercise and all alternatives with zero bids were omitted. Restricting the analysis to individuals' consideration sets allows for a more meaningful comparison and eliminates unnecessary noise in the analysis.

The seven fruit products that were included in the experiment were: 1) California Wonderful fresh pomegranate (the predominant variety in the market); 2) Texas Red fresh pomegranate; 3) Texas Salavatsky fresh pomegranate; 4) Ready-to-eat California pomegranate arils; 5) Ready-to-eat Texas pomegranate arils; 6) Pomegranate juice; and 7) a pineapple which served as the control.

In order to avoid biases in the results from confusions or misunderstanding, the participants were presented with extensive instructions about the experiment and the mechanisms involved. They were also informed that they would have to pay for any purchases made during the experiment, but that they could not have more than one purchase. Furthermore, two practice rounds were conducted, using soft drinks and snacks, before the experiments in order to train the participants. The market prices were posted and discussed after each practice round to make sure that everyone was familiar with the process and how it works.

Following the practice rounds, the participants engaged in four different rounds of auctions and ranking. In each round, the subjects were required to submit bids and rankings for the alternatives. The first round served as a baseline round where the subjects were given no information about the products. After the baseline round, the subjects participated in three treatment rounds and were asked to submit their responses after each treatment. The three treatment rounds consisted of a tasting treatment and two informational treatments concerning health benefits and anti-cancer properties. In the tasting treatment, the subjects were given the opportunity of tasting a small sample of each product, around 2 oz., before submitting their responses. Tasting was completely voluntary, but none of the participants refused to taste any of the fruit products. For the health information treatment, participants were provided with some health and nutritional benefits of each product before submitting their responses. Finally, during the anti-cancer information treatment, the subjects were presented with information on the potential anti-cancer properties of pomegranate products before submitting their bids and rankings. The order of the treatment rounds was randomized across sessions so as to minimize any noise from ordering effects. The market price and buyers were announced at the end of the experiment, after all treatments were completed. The market prices were not revealed after each round to control for any confounding effects and bid affiliations and to avoid influencing the ranking decisions of participants by knowing the prices. Only one round and one product were randomly selected as binding. Everyone who bid higher than the market price for the specified product in the binding round ended up purchasing that product for the market price.

4. Methodology:

Various versions of the logit model were utilized to produce different tests and parameter estimates, which were analyzed and discussed to provide a better understanding of the consumers' decision-making process and the forces that stimulate inconsistencies in their preferences. This section includes a brief overview over the econometric models and concepts that were used for data analysis.

4.1 Data Explosion process

The ranked-ordered logit model is commonly used to estimate preferences over the whole set of alternatives. This is done by exploding the ranking data into a series of choice decisions, where the chosen alternative is removed from each subsequent choice decision. First, the most preferred alternative is chosen from the initial set. Next, the second most preferred alternative is chosen from the remaining set of alternatives, which does not include the most preferred alternative. The process continues until all the rankings are accounted for. Since revealing the second least preferred alternative also reveals the least preferred, an initial set with J alternatives would be exploded into $J - 1$ choice decisions.

The two value elicitation methods have different natures in the sense that the auction experiment allows for ties among certain alternatives (one could bid the same value for two or more alternatives), while the ranking procedure requires the participants to break the ties by assigning different ranks to the alternatives. In order to account for this difference, ties in the auction experiment were split and a bootstrap procedure was used to assign the rankings. One thousand random draws were taken for each alternative to determine the assignment of the highest rank.

4.2 The Conditional Logit Model

In our regression model, we follow McFadden's random utility theory (McFadden 1974). In this framework, the utility of a consumer $n \in \{1,2,\dots,N\}$ for alternative $j \in \{1,2,\dots,J\}$ in time period or treatment $t \in \{1,2,\dots,T\}$ is stochastic and depends on the attributes of the particular alternative (Train 2009). This utility is represented by the function

$$U_{njt} = V_{njt} + \varepsilon_{njt} \qquad (1)$$

Where V_{njt} is the deterministic component, which depends on the attributes of the alternative X_{njt}. The relationship is commonly linear and takes the form $V_{njt} = \beta' X_{njt}$. The stochastic component ε_{njt} represents the error term, which is usually assumed to be independently and identically distributed (iid) extreme value.

Each individual picks the alternative that maximizes his/her utility. This means that alternative j will be chosen by individual n at time period or treatment t if and only if $U_{njt} > U_{nkt} \ \forall\, k \neq j$ which implies that $V_{njt} + \varepsilon_{njt} > V_{nkt} + \varepsilon_{nkt} \ \forall\, k \neq j$ or $\varepsilon_{nkt} < \varepsilon_{njt} + V_{njt} - V_{nkt} \ \forall\, k \neq j$. Under the iid extreme value assumption, the probability of choosing alternative j is given by:

$$P_{njt} = \frac{e^{\beta' X_{njt}}}{\Sigma_k e^{\beta' X_{nkt}}} \tag{2}$$

which are the logit probabilities. If we estimate a ranked-ordered logit, or manually explode the data and estimate a conditional logit, we can calculate the probability of a particular ranking instead of just the probability of choosing a certain alternative. This would allow us to elicit preferences over the whole set of alternatives rather than just the top ranked alternative. For illustration purposes, let us assume that the individual reported the following ranking for the alternatives: $alt(1) \succ alt(2) \succ alt(3) \succ \cdots \succ alt(J)$. Since the data is exploded into a series of $J-1$ choice decisions, the conditional logit model calculates the probability of this ranking as the product of the probabilities of each chosen alternative at each choice decision. Thus, the probability of the ranking is given by:

$$Prob\left[U_{n1t} > U_{n2t} > \cdots > U_{nJt}\right] = \frac{e^{\beta' X_{n1t}}}{\Sigma_{j=1}^{J} e^{\beta' X_{njt}}} \cdot \frac{e^{\beta' X_{n2t}}}{\Sigma_{j=2}^{J} e^{\beta' X_{njt}}} \cdot \cdots \cdot \frac{e^{\beta' X_{n(J-1)t}}}{\Sigma_{j=J-1}^{J} e^{\beta' X_{njt}}} \tag{3}$$

4.3 Conditional Logit with Partial Rankings:

In order to present a more detailed analysis of preference inconsistencies and provide a deeper understanding of the consumer decision-making process, we have used various partial rankings in our study. The conditional logit model was utilized to run top and bottom partial rankings. The top rankings included Trank1, Trank2, Trank3, Trank4, Trank5, and Trank6 which respectively stand for the best alternative, best two alternatives, all the way until the best 6 alternatives (which is basically the full ranking). On the other hand, the bottom rankings included the worst alternative (Brank1), the worst two alternatives (Brank2), the worst three alternatives (Brank3), and the worst four alternatives (Brank4). The partial rankings were estimated on ranks and bids (recoded as implied ranks).

When applying the conditional logit model on partial rankings, only the relevant data are included based on the specified ranking and everything else is dropped out of the data set. For example, if we wish to consider the top two alternatives only, then after exploding the data, only the first two choice decisions are user. With this reasoning, it is straightforward to derive the probabilities of the different partial rankings. Following our previous example, where the individual reports the ranking $alt(1) \succ alt(2) \succ alt(3) \succ \cdots \succ alt(J)$, if we only select the best alternative then the ranking probability would simply be the probability of choosing the top alternative. This is given by the following equation:

$$Prob\left[U_{n1t} > U_{n2t} > \cdots > U_{nJt}\right] = \frac{e^{\beta' X_{n1t}}}{\Sigma_{j=1}^{J} e^{\beta' X_{njt}}} \tag{4}$$

If we choose to select the top two alternatives then the ranking probability would be:

$$Prob\left[U_{n1t} > U_{n2t} > \cdots > U_{nJt}\right] = \frac{e^{\beta' X_{n1t}}}{\Sigma_{j=1}^{J} e^{\beta' X_{njt}}} \cdot \frac{e^{\beta' X_{n2t}}}{\Sigma_{j=2}^{J} e^{\beta' X_{njt}}} \tag{5}$$

The ranking probabilities of the other partial ranks can be derived in a similar fashion.

4.4 Ranking Vs. Implied Ranking: Structural Stability Tests:

The stability of parameters was tested using two different methods, which were used to determine whether the participants exhibited preference inconsistencies between the ranking procedure and the implied rankings of the auction experiment. By examining structural changes in parameters, those tests would help assess whether the subjects were following the same behavioral rules under the two preference elicitation mechanisms. First, a likelihood ratio test of the following form was completed: $LR = -2[L_{r+ir} - (L_r + L_{ir})]$ where L_r stands for the likelihood function based on the ranking data, L_{ir} is the likelihood function based on the implied ranking data, and L_{r+ir} represents the likelihood function based on the pooled data (ranking and implied ranking). Rejecting this test would imply significant structural changes in the parameters between the auction and ranking. This means that the preferences were not stable and consumers were using different rules to value the products across the two valuation mechanisms.

Another test of parameter stability was proposed by Allison and Christakis (1994). In this method, testing for structural stability in the parameters is performed by estimating stage or preference elicitation-specific covariates. This was done by creating a mechanism specific indicator variable that interacts with each parameter. A chi-squared test for the joint significance of one of the preference elicitation-specific covariates is conducted. Failure to reject this test would indicate stability in the parameters, while rejecting the test implies instability and inconsistency in preferences.

4.5 Ranking Vs. Implied Ranking: Parameter Symmetry Tests:

The parameter symmetry test is used to measure differences in unobserved error variances across the partial rankings and implied rankings. A heteroscedastic conditional logit model was used to estimate a scale parameter σ associated with the ranking data. The scale parameter was then tested for significance to determine whether there was symmetry in the parameters between ranking and implied ranking. A significant scale parameter would imply asymmetry and hence preference inconsistency between the implied ranking of the auction experiment and the rankings procedure.

4.6 Ranking Vs. Implied Ranking: systematic differences in preferences:

Systematic differences between the valuation methods could occur when the subjects deliberately change their behavior based on the mechanism or situation they are facing. It is necessary to check for the existence of systematic differences between the two value elicitation methods in our case since the subjects experience different environments under each. Subjects make their decisions independently under the ranking procedure, while the auction exercise requires interaction between the subjects in the sense that their decisions will affect the market price. In this study, we use the seemingly unrelated estimation test (SUEST) to check for systematic differences between the two elicitation methods. This test uses a similar but more general approach than the Hausman specification. It combines the parameter estimates and covariance matrices of both models under one parameter vector and individually tests for any relationship between the equivalent parameters. Rejecting the test implies systematic differences between the models.

4.7 Ranking Vs. Implied Ranking: Predictive Power Tests:

Since the number of observations varies across the partial ranking and implied ranking models, comparing them using a likelihood-based goodness of fit would be inappropriate. Instead, we can compare the predictive power of the models. Tjur (2009) introduced a goodness of fit measure for logistic regression models. This measure is bounded between zero and one and is equivalent to the standard R^2 used in linear regressions. It is calculated by splitting the dependent variable into two categories (events and nonevents). Then the mean predicted probability is calculated for events ($y = 1$) and nonevents ($y = 0$) and the difference between those two means is taken as Tjur's R^2. Intuitively, if a model makes good predictions then the cases with events should have higher mean predictions than the cases with nonevents. Hence the higher the value of Tjur's R^2

the more powerful the model is in predicting choices. Since this measure is not based on a likelihood function, it can be applied to compare non-nested models.

5. Results and Discussion

Table 2 and table 3 show the parameter estimates of the conditional logit model for the ranking and implied ranking data respectively. Each of the partial ranking models was applied to both types of data in order to provide a detailed comparison of preferences between the auction experiment and ranking procedure. The models used product attributes as independent variables. The attributes considered were split into two categories. The first category (Variety) included Texas Red and Texas Salavatski, while the second category (Product Form) was associated with the presentation of the product and included ready-to-eat, juice, and pineapple. As shown in the tables, the number of observations increases as more partial rankings are considered. This is evident in the ranking data, where the number of observations goes from 895 in Trank1 to 3139 in Trank6 and in the implied ranking data, where the number of observations goes from 1127 in Timprank1 to 4504 in Timprank6. This result is straightforward and is based on the process of exploding the data. The rankings, and implied rankings, were exploded into different choice decisions. The most preferred alternative was chosen in the first choice decision, the second most preferred alternative was chosen in the second choice decision and so on until the full ranking was revealed. This means that including more partial rankings is equivalent to including more choice decisions and more observations. The number of observations in the implied ranking models was consistently higher than that in the equivalent ranking models. This difference is attributable, at least in part, to the way individuals defined their consideration sets under each value elicitation mechanism. For example, some individuals included a certain alternative in their consideration set under the auction experiment but not under the ranking procedure. This in turn resulted in a smaller consideration set for the ranking procedure, a lower number of choice decisions in the exploded data, and a lower number of observations under the ranking model.

The ranking and implied ranking data were pooled and regressed under each partial ranking in order to conduct the likelihood ratio test for parameter stability. The test was rejected for all partial rankings (top and bottom). This implies that the parameters were not stable across the two elicitation methods, which is a signal of preference inconsistency. The behavioral rules used by participants to value the alternatives under the auction/bid were different than those used under the ranking exercise. The statistical significance of the test increases dramatically as more partial rankings are included. The value of the test statistic is smallest for the Trank1, Brank1 and Brank2 models. This suggests that preferences are most stable for the highest and lowest ranked alternatives, while the stability drastically decreases for the middle ranked alternatives. Although consumers follow different sets of rules under the two valuation mechanisms, the rules are more affiliated for the top and bottom partial rankings than they are for the middle rankings. Intuitively speaking, ranking the middle alternatives is more difficult for the individual than ranking the top and bottom alternatives. This is because extreme like and dislike are more obvious to decision-makers than the moderate, middle-ground feelings they experience with the middle ranked alternatives. Due to the simplicity of the decision, the consumers do not exhibit as many preference inconsistencies in the top and bottom rankings as they do in the middle rankings.

Table 4 shows results for parameter stability using the method introduced by Allison and Christakis (1994). The results suggest another aspect of the decision-making process. This method estimates two sets of covariates, one for each preference elicitation mechanism. A chi-squared test for the joint significance of the parameter estimates under the implied ranking (auction) data was rejected for all partial ranking models except Brank2 and Brank4. In other words, the parameters were not significant for those two partial ranking models. This confirms the fact that subjects might be using similar strategies for ranking the worst alternatives. Perhaps the subject invests effort in accurately ranking the top alternatives then ranks the bottom alternatives in a random fashion due to a lack of interest in them. This randomness in turn could result in similar patterns when taken over a large sample of observations, which might increase the affiliation of preferences for those rankings. Following this reasoning, one might expect an equal chance of observing symmetry or asymmetry for the bottom ranking models and the fact that it was observed in two out of the four bottom ranking models strengthens this hypothesis. Moreover, this conclusion is also affirmed by the results in table 2 and table 3, where the statistical significance of the likelihood ratio test is generally lower for the bottom partial ranking models.

A heteroscedastic conditional logit model was estimated on the partial rankings and partial implied rankings to test the symmetry in preferences between the two valuation methods. As shown in table 5, a scale parameter was estimated to assess the differences in the variance of parameters between the auction experiment and ranking procedure. This was done by attaching the scale parameter to the variance of the coefficients associated with the ranking data. The higher that parameter, the lower the variance of the coefficients in the ranking data compared to the implied ranking data and vice versa. The scale parameter was negative for all the top partial rankings and the full ranking. This negative sign suggests that the variance of the coefficients was higher in the ranking compared to the implied ranking models. However, the scale parameter was only significant for the Trank2, Trank3, and Trank4 models, which implies strong asymmetry in the parameters under those partial ranking models. Also, the absolute value of the scale parameter increased from .303 in Trank2 to .625 in Trank3 to .862 in Trank4 indicating an increase in asymmetry. This result confirms the findings in parameter stability. Since the decision-maker's job gets more difficult as he considers the middle ranked alternatives, he may be less accurate and less consistent in evaluating those alternatives. Hence, his preferences appear to be more inconsistent when those alternatives are included in the model. Furthermore, the fact that the scale parameter was insignificant for all of the bottom partial rankings is in line with the conclusion that subjects' disinterest with the least preferred alternatives drives them to adopt random valuations for those alternatives, which results in more symmetry in preferences when taken over many observations. It also explains why the significance of the scale parameter decreases radically in Trank5 and Trank6, when those bottom rankings are included.

Results from the seemingly unrelated estimation test are presented in table 2 and table 3. Here, each of the top and bottom partial ranking models was compared with the respective partial implied ranking model in order to determine if and where systematic differences exist between the two value elicitation mechanisms. Besides the top rank model (Trank1, Timprank1), the test was rejected for all other top partial ranking models which implies that systematic differences between the valuation methods were evident in those models. This result is quite interesting since it adds perspective to the previous findings. It suggests that although individuals use different behavioral rules under the two elicitation mechanisms, they still arrive at similar decisions concerning their most preferred alternative, at least systematically. However, the same cannot be said about the other alternatives in their consideration sets since the models for those alternatives differ systematically and their parameters are asymmetric and structurally unstable. Furthermore, the statistical significance of the test increases substantially when the middle rankings are considered indicating higher systematic differences between the elicitation mechanisms. Our previous conclusion concerning the bottom ranked alternatives is reinforced by the seemingly unrelated estimation test. The statistical significance of the test is somewhat lower for the bottom partial ranking models and we actually failed to reject the test for the bottom 3 ranking model (Brank3, Bimprank3). This presents more evidence that subjects evaluate the worst rankings in a random manner which might result in similar patterns when taken over many observations. More importantly, this conclusion and the fact that the test was rejected for the worst ranking model (Brank1, Bimprank1) suggest that considering the top ranked alternative alone is a more accurate approach across the two elicitation mechanisms.

The Calculated Tjur's R^2 in the conditional logit model in table 2 and table 3 indicate that the implied ranking from the auction/bid experiment produced better predictions than the ranking procedure. Tjur's R^2 was significantly higher for the implied ranking data than for the ranking data in all of the top partial ranking and the full ranking models. This may be a reflection of the incentive compatibility and accuracy of the revealed preference elicitation mechanism. In addition, it is also possible that people are more accustomed to placing values on products than they are to ranking them. An individual is faced with the need to value products on a daily basis. It is how he decides whether to purchase something or not. If it is priced less than his valuation he would purchase it, otherwise he would leave it on the shelf. However, people seldom encounter situations where they have to rank the relative attractiveness of several products. The unfamiliarity with this exercise increases the cognitive effort associated with it. Thus, in an effort to avoid this cognitive cost, the individual is pushed to make a hasty or inaccurate decision regarding the ranking. This increases the randomness and decreases the reliability in his reported values, which decreases the predictive power of the model. This conclusion is supported by the results from the parameter symmetry test in table 4, where the estimated coefficients in the ranking data had a higher variance than those in the implied ranking data. The value of Tjur's R^2

also decreased as more partial rankings were added. As shown in table 2 and table 3, it ranged from 0.257 in Trank1 to 0.089 in Trank6 for the ranking data, while for the implied ranking data it went from 0.261 in Timprank1 all the way to 0.107 in Timprank6. This result is expected since preferences are less certain concerning the middle ranked alternatives, which means that there is a high cognitive effort involved in valuing them. Hence, as more partial rankings are considered, which include those middle rankings, the predictive power of the model should decrease. The bottom ranking models showed more ambiguous results regarding Tjur's R^2. The differences were less consistent between the ranking data and the implied ranking data. The predictive power was actually higher for the ranking data than for the implied ranking data in the bottom 2, bottom 3 and bottom 4 partial ranking models. The lack of interest in those alternatives caused individuals to treat them indifferently in the auction and the ranking, since they evaluated them less carefully. This fact buffered the effect of unfamiliarity with the ranking exercise, which tipped the Tjur's R^2 measure in its favor.

6. Summary and Conclusion:

The controversy over the validity of stated preference methods in eliciting true valuations has encouraged more comparisons of consumer preferences between stated and revealed elicitation mechanisms. This in turn sparked more analyses of preference inconsistencies and the factors that cause individuals to adopt different behavioral rules under different valuation mechanisms. This study utilized an extensive comparison between a non-hypothetical ranking procedure and a non-hypothetical, incentive compatible auction experiment in order to bolster our understanding of the consumer's decision making process, provide a robust comparison between stated and revealed preferences, and explain the main forces that give rise to preference inconsistencies. While many previous comparisons were centered on willingness-to-pay estimates, our analysis used ordinal data by recoding the bids from the auction experiment into implied rankings. This was done to allow for a more accurate comparison.

The results of two tests for structural stability indicated that the parameters were not stable between the preference elicitation mechanisms for any of the partial ranking models. This implied that consumers were using different behavioral rules to evaluate the alternatives under the auction and ranking exercises. However, the affiliation was higher for the top ranked alternatives since preferences are more certain about this extreme. Moreover, tests for parameter symmetry indicated that parameters were more asymmetric when the middle rankings are included since preferences were vague for those alternatives and decisions were more complicated over them. On the other hand, the stability and symmetry of the parameters in the bottom partial ranking models point at the conclusion that individuals assign random valuations for those alternatives due to a lack of interest in them.

Besides the best ranked alternative, individuals approached the alternatives under the two value elicitation methods in a systematically different manner. This result indicated that even though the participants had different behavioral rules under the two valuation mechanisms, they still arrived at similar decisions concerning their most preferred alternative. In contrast, their random valuation of the worst alternatives resulted in random patterns that were similar in some cases as was evident in the worst three alternatives model (Brank3, Bimprank3) where the seemingly unrelated estimation test for systematic differences was not rejected.

Overall, the revealed preference auction data was more reliable than the ranking data and produced more accurate predictions. The implied ranking models had a higher predictive power than the ranking models based on Tjur's R^2. Furthermore, the variance of parameter estimates was lower for the implied ranking data. The incentive compatibility of experimental auctions did make a difference in forming prediction orderings and their variance. It is possible that the behavioral process was more unfamiliar in the ranking exercise, which caused individuals to approach it less accurately. The predictive power decreased, in general, as more partial rankings were considered. This result supported the previous findings that decisions over top and bottom ranked alternatives are easier than decisions over middle ranked alternatives.

In conclusion, this article provided strong results concerning preference inconsistencies between auction exercises and ranking procedures. Compared to the dominant and more accurate auction exercise, ranking procedures perform fairly well in eliciting the top ranked alternative. However, the accuracy of the ranking exercise in eliciting true valuations substantially decreases when the other partial rankings are included as is evident by the drastic decrease in the predictive power of the models based on the ranking exercise compared to those of the auction mechanism and the fact that subjects' responses differ systematically under both mechanisms when partial rankings are considered. Hence, there is evidence in favor of the validity of stated preference mechanisms but only when eliciting preferences over the best alternative.

7. References:

- Abramovitz, Moses, Maurice Allais, Roy George Douglas Allen, Oskar Nikolayevich Anderson, Giovanni Battista Antonelli, St Thomas Aquinas, Heinz Wolfgang Arndt et al. "The New Palgrave Dictionary of Economics, /List of articles Contributor."
- Allison, Paul D., and Nicholas A. Christakis. "Logit models for sets of ranked items." Sociological methodology 24, no. 1994 (1994): 199-228.
- Alwin, Duane F., and Jon A. Krosnick. "The measurement of values in surveys: A comparison of ratings and rankings." Public Opinion Quarterly 49, no. 4 (1985): 535-552.
- Berg, Joyce E., John W. Dickhaut, and John R. O'Brien. "Preference reversal and arbitrage." Research in experimental economics 3 (1985): 31-72.
- Böckenholt, Ulf. "Thurstonian representation for partial ranking data." British Journal of Mathematical and Statistical Psychology 45, no. 1 (1992): 31-49.

- Boyle, Kevin J., Thomas P. Holmes, Mario F. Teisl, and Brian Roe. "A comparison of conjoint analysis response formats." American Journal of Agricultural Economics 83, no. 2 (2001): 441-454.
- Bunch, David S., Mark Bradley, Thomas F. Golob, Ryuichi Kitamura, and Gareth P. Occhiuzzo. "Demand for clean-fuel vehicles in California: a discrete-choice stated preference pilot project." Transportation Research Part A: Policy and Practice 27, no. 3 (1993): 237-253.
- Caparrós, Alejandro, José L. Oviedo, and Pablo Campos. "Would you choose your preferred option? Comparing choice and recoded ranking experiments."American Journal of Agricultural Economics 90, no. 3 (2008): 843-855.
- Carpenter, Jason M., and Marguerite Moore. "Consumer demographics, store attributes, and retail format choice in the US grocery market." International Journal of Retail & Distribution Management 34, no. 6 (2006): 434-452.
- Chu, Yun-Peng, and Ruey-Ling Chu. "The subsidence of preference reversals in simplified and marketlike experimental settings: A note." The American Economic Review (1990): 902-911.
- Corrigan, Jay R., Dinah Pura T. Depositario, Rodolfo M. Nayga, Ximing Wu, and Tiffany P. Laude. "Comparing open-ended choice experiments and experimental auctions: An application to golden rice." American Journal of Agricultural Economics 91, no. 3 (2009): 837-853.
- Goldstein, William M., and Hillel J. Einhorn. "Expression theory and the preference reversal phenomena." Psychological review 94, no. 2 (1987): 236.
- Grether, David M., and Charles R. Plott. "Economic theory of choice and the preference reversal phenomenon." The American Economic Review (1979): 623-638.
- Hanley, Nick, Robert E. Wright, and Vic Adamowicz. "Using choice experiments to value the environment." Environmental and resource economics11, no. 3-4 (1998): 413-428.
- Harzing, Anne-Wil, Joyce Baldueza, Wilhelm Barner-Rasmussen, Cordula Barzantny, Anne Canabal, Anabella Davila, Alvaro Espejo et al. "Rating versus ranking: What is the best way to reduce response and language bias in cross-national research?." International Business Review 18, no. 4 (2009): 417-432.
- Hensher, David A. "Stated preference analysis of travel choices: the state of practice." Transportation 21, no. 2 (1994): 107-133.
- Hensher, David A., and Chinh Ho. "Identifying a behaviourally relevant choice set from stated choice data." Transportation: 1-21.
- Kagel, John H., and Dan Levin. "Independent private value auctions: Bidder behaviour in first-, second-and third-price auctions with varying numbers of bidders." The Economic Journal (1993): 868-879.
- Katok, Elena, and Alvin E. Roth. "Auctions of homogeneous goods with increasing returns: experimental comparison of alternative “Dutch” auctions."Management Science 50, no. 8 (2004): 1044-1063.
- Kim, Betty E., Darryl Seligman, and Joseph W. Kable. "Preference reversals in decision making under risk are accompanied by changes in attention to different attributes." Frontiers in neuroscience 6 (2012).
- Lichtenstein, Sarah, and Paul Slovic. "Response-induced reversals of preference in gambling: An extended replication in Las Vegas." Journal of Experimental Psychology 101, no. 1 (1973): 16.

- List, John A., and Craig A. Gallet. "What experimental protocol influence disparities between actual and hypothetical stated values?." Environmental and Resource Economics 20, no. 3 (2001): 241-254.
- Loomes, Graham, and Robert Sugden. "Regret theory and measurable utility."Economics Letters 12, no. 1 (1983): 19-21.
- Lusk, Jayson L. "Using experimental auctions for marketing applications: a discussion." Journal of Agricultural and Applied Economics 35, no. 02 (2003): 349-360.
- Lusk, Jayson L., and Ted C. Schroeder. "Auction bids and shopping choices."Advances in Economic Analysis & Policy 6, no. 1 (2006).
- Mackenzie, John. "A comparison of contingent preference models." American Journal of Agricultural Economics 75, no. 3 (1993): 593-603.
- McAdams, Callie, Marco A. Palma, Charles Hall, and Ariun Ishdorj. "A Nonhypothetical Ranking and Auction Mechanism for Novel Products." Journal of Agricultural and Applied Economics 45, no. 01 (2013).
- McFadden, Daniel. "The measurement of urban travel demand." Journal of public economics 3, no. 4 (1974): 303-328.
- McNeil, Barbara J., Stephen G. Pauker, Harold C. Sox Jr, and Amos Tversky. "On the elicitation of preferences for alternative therapies." New England journal of medicine 306, no. 21 (1982): 1259-1262.
- Mellers, Barbara A., Shi-jie Chang, Michael H. Birnbaum, and Lisa D. Ordonez. "Preferences, prices, and ratings in risky decision making." Journal of Experimental Psychology: Human Perception and Performance 18, no. 2 (1992): 347.
- Morrison, Mark, and Kevin Boyle. Comparative reliability of rank and choice data in stated preference models. Faculty of Commerce, Charles Sturt University, 2001.
- Mowen, John C., and James W. Gentry. "Investigation of the preference-reversal phenomenon in a new product introduction task." Journal of Applied Psychology 65, no. 6 (1980): 715.
- Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. "A meta-analysis of hypothetical bias in stated preference valuation." Environmental and Resource Economics 30, no. 3 (2005): 313-325.
- Pavan, M., and R. Todeschini. "New indices for analysing partial ranking diagrams." Analytica chimica acta 515, no. 1 (2004): 167-181.
- Revelt, David, and Kenneth Train. "Mixed logit with repeated choices: households' choices of appliance efficiency level." Review of economics and statistics 80, no. 4 (1998): 647-657.
- Rutström, E. Elisabet. "Home-grown values and incentive compatible auction design." International Journal of Game Theory 27, no. 3 (1998): 427-441.
- Sayadi, Samir, M. Carmen Gonzalez Roa, and Javier Calatrava Requena. "Ranking versus scale rating in conjoint analysis: Evaluating landscapes in mountainous regions in southeastern Spain." Ecological Economics 55, no. 4 (2005): 539-550.
- Shogren, Jason F., Michael Margolis, Cannon Koo, and John A. List. "A random nth-price auction." Journal of economic behavior & organization 46, no. 4 (2001): 409-421.
- Siikamäki, Juha, and David F. Layton. "Discrete choice survey experiments: a comparison using flexible methods." Journal of Environmental Economics and Management 53, no. 1 (2007): 122-139.

- Su, Lianfan, Brian D. Adam, Jayson L. Lusk, and Frank Arthur. "A comparison of auction and choice experiment: An application to consumer willingness to pay for rice with improved storage management." In 2011 Annual Meeting, July 24-26, 2011, Pittsburgh, Pennsylvania, no. 103975. Agricultural and Applied Economics Association, 2011.
- Tjur, Tue. "Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination." The American Statistician 63, no. 4 (2009): 366-372.
- Train, Kenneth E. 2009. Discrete choice methods with simulation: Cambridge university press
- Tversky, Amos, Shmuel Sattath, and Paul Slovic. "Contingent weighting in judgment and choice." Psychological review 95, no. 3 (1988): 371.
- Tversky, Amos, Paul Slovic, and Daniel Kahneman. "The causes of preference reversal." The American Economic Review (1990): 204-217.
- Vickrey, William. "Counterspeculation, auctions, and competitive sealed tenders." The Journal of finance 16, no. 1 (1961): 8-37.
- Wardman, Mark. "A comparison of revealed preference and stated preference models of travel behaviour." Journal of Transport Economics and Policy (1988): 71-91.
- Zeiler, Kathryn, and Charles R. Plott. "The willingness to pay/willingness to accept gap, the endowment effect, subject misconceptions and experimental procedures for eliciting valuations." American Economic Review (2004).

Table1: Demographic and Behavioral Characteristics of Shoppers

Variable	Category	Sample		
		Mean	Std. Dev.	Percent
Age (years)		42.84	17.51	
	Under 29			34.83%
	30-39			11.94%
	40-49			14.43%
	50-59			21.89%
	60-69			7.46%
	70 and over			9.45%
Household Size (Individuals)		2.24	1.15	
Education	High School Diploma or Less			11.44%
	Bachelor's Degree or at least some College			0.61
	Graduate Courses or more			27.86%
Gender	Female			68.66%
	Male			31.34%
Marital Status	Married			54.23%
	Not Married			45.77%
annual Household Income ($)		53,693	36,973	
Primary Shopper	Primary Shopper			88%
	Secondary Shopper			12%
Household Spending on Food ($/week)		109.13	75.49	
Household Spending on Fruits and Vegetables ($/weeks)		25.13	17.72	
Fruits and Vegetables on Hand (lbs.)		6.37	4.65	
Have a Serious Health Issues	Yes			28.50%
	No			71.50%
Tobacco Use (% of days per year)	Yes	20.79	57.77	
Exercise (% of days per year)		43.52	38.97	

Table 2. Conditional Logit Parameter Estimates of the Exploded Data for the Ranking

	Variable		Trank1 [a]	Trank2	Trank3	Trank4	Trank5	Trank6	Brank1	Brank2	Brank3	Brank4
Variety												
	Texas Red		-1.0329 *	-0.2715	0.1134	-0.1156	-0.1951	-0.1196	0.0611	0.3459	-0.2128	-0.3357 *
		se	0.5661	0.3054	0.2165	0.1856	0.1571	0.1418	0.2821	0.3980	0.2410	0.1929
	Texas Salavatski		-0.5903 *	-0.0888	0.1678	0.0906	0.0219	0.0438	-0.2222	0.2737	-0.1179	-0.1139
		se	0.3445	0.1976	0.1431	0.1216	0.1091	0.1033	0.2359	0.3686	0.2108	0.1556
Product Form												
	Ready to Eat		0.0063	0.2879	0.4272 ***	0.4076 ***	0.2904 **	0.3177 ***	-0.4258 *	0.4699	-0.4499	0.2015
		se	0.3403	0.2070	0.1532	0.1291	0.1197	0.1153	0.2496	0.6514	0.3231	0.1995
	Juice		0.7991 **	1.1037 ***	0.9359 ***	0.7173 ***	0.4843 ***	0.3643 **	0.1585	-2.9900 **	-1.8451 ***	-0.8875 ***
		se	0.3381	0.2258	0.1910	0.1675	0.1553	0.1524	0.2652	1.1568	0.4373	0.2891
	Pinneaple		1.8430 ***	1.8465 ***	1.6404 ***	1.4040 ***	1.1483 ***	1.1296 ***	-0.9020 ***	-1.5733	-1.0224 **	-0.2309
		se	0.3020	0.2235	0.1893	0.1675	0.1563	0.1519	0.3281	1.1505	0.4444	0.3028
	NOBS [b]		895	1618	2192	2625	2937	3139	895	202	514	947
	Tjur's R sq [c]		0.2570	0.1624	0.0957	0.0825	0.0714	0.0888	0.0724	0.2086	0.1097	0.0780
	Pseudo R sq		0.2513	0.1454	0.0737	0.0576	0.0400	0.0352	0.0387	0.1822	0.0640	0.0289
	LR-Test of Pref Stability [d]		43.9384	142.1561	317.2361	503.5535	731.0338	999.0810	66.7626	39.6854	83.9758	201.9930
	SUEST [e]		9.1500	14.2200	22.7000	22.1400	18.4000	21.7300	13.9200	13.3400	4.8000	11.8200
	P-Value		0.1032	0.0143	0.0004	0.0005	0.0025	0.0006	0.0161	0.0204	0.4405	0.0373

(a) Single (*), double (**), and triple (***) asterisks are used to denote significance at the 0.10, 0.50, and 0.01, respectively. (Std Errors)

(b) number of observations of the exploded data

(c) Tjur's coefficient of descrimination to measure goodness of fit (predictive power) of model. It is independent of number of observations

(d) LR test for parameter stability. Three different models were used: model using ranking data, model using implied ranking data, and model using pooled data from ranking and implied ranking

(e) seemingly unrelated estimation test for systematic differences between the models. This is a generalized form of the Hausman test.

Table 3. Conditional Logit Parameter Estimates of the Exploded Data for the Implied Ranking

	Variable		Timprank1 [(a)]	Timprank2	Timprank3	Timprank4	Timprank5	Timprank6	Bimprank1	Bimprank2	Bimprank3	Bimprank4
Variety												
	Texas Red		0.3234	-0.2355	-0.1809	-0.0737	-0.0548	-0.1081	0.1848	-0.1708	-0.1007	-0.0389
		se	0.5732	0.3719	0.2438	0.1878	0.1424	0.1271	0.2184	0.2899	0.1797	0.1531
	Texas Salavatski		-0.5118	-0.1801	-0.0962	0.0277	0.0157	0.0730	-0.3395 *	0.3938	0.1649	0.2043 *
		se	0.3653	0.2064	0.1352	0.1130	0.0985	0.0906	0.2000	0.2396	0.1537	0.1230
Product Form												
	Ready to Eat		1.1726 ***	0.8810 ***	0.9597 ***	0.8794 ***	0.6702 ***	0.6081 ***	-0.57309 ***	0.1341	-0.2856	0.3172 **
		se	0.4290	0.2319	0.1543	0.1248	0.1090	0.1024	0.210392	0.3737	0.2319	0.1536
	Juice		2.2937 ***	2.0967 ***	1.7120 ***	1.6260 ***	1.3034 ***	1.2297 ***	-1.13697 ***	0.3508	-0.5986 *	0.4093 *
		se	0.4280	0.2413	0.1819	0.1556	0.1395	0.1325	0.299331	0.4816	0.3544	0.2304
	Pinneaple		2.8993 ***	2.5096 ***	2.2229 ***	2.0928 ***	1.7759 ***	1.7200 ***	-1.94204 ***	0.8554	-0.5304	0.2390
		se	0.4179	0.2414	0.1807	0.1577	0.1421	0.1358	0.3645	0.5925	0.4443	0.3055
	NOBS [(b)]		1127	2102	2915	3598	4119	4504	1127	385	906	1589
	Tjur's R sq [(c)]		0.2606	0.2301	0.1537	0.1338	0.1010	0.1065	0.0988	0.0942	0.0595	0.0526
	Pseudo R sq		0.2908	0.2484	0.1647	0.1365	0.0928	0.0830	0.0865	0.0281	0.0117	0.0102
	LR-Test of Pref Stability [(d)]		43.9384	142.1561	317.2361	503.5535	731.0338	999.0810	66.7626	39.6854	83.9758	201.9930
	SUEST [(e)]		9.1500	14.2200	22.7000	22.1400	18.4000	21.7300	13.9200	13.3400	4.8000	11.8200
	P-Value		0.1032	0.0143	0.0004	0.0005	0.0025	0.0006	0.0161	0.0204	0.4405	0.0373

(a) Single (*), double (**), and triple (***) asterisks are used to denote significance at the 0.10, 0.50, and 0.01, respectively. (Std Errors)

(b) number of observations of the exploded data

(c) Tjur's coefficient of descrimination to measure goodness of fit (predictive power) of model. It is independent of number of observations

(d) LR test for parameter stability. Three different models were used: model using ranking data, model using implied ranking data, and model using pooled data from ranking and implied ranking

(e) seemingly unrelated estimation test for systematic differences between the models. This is a generalized form of the Hausman test.

Table 4. Conditional Logit Parameter Estimates of the Exploded Data with Implied Ranking Taken as a Treatment

	Variable		Trank1 [(a)]	Trank2	Trank3	Trank4	Trank5	Trank6	Brank1	Brank2	Brank3	Brank4
Variety												
	RTexas Red		-0.72839	-0.01402	0.296649	0.076617	-0.03576	0.0942	-0.02501	0.618486 *	0.028706	-0.08406
		se	0.557138	0.289807	0.198988	0.172548	0.145879	0.129721	0.264634	0.371358	0.221951	0.179655
	RTexas Salavatski		-0.41637	0.05129	0.271935 **	0.210107 *	0.135718	0.185119 *	-0.28504	0.500109	0.082486	0.085172
		se	0.343826	0.193701	0.137964	0.117574	0.105114	0.098183	0.225855	0.344264	0.197065	0.147227
Product Form												
	RReady to Eat		0.28435	0.511145 ***	0.599897 ***	0.579945 ***	0.399039 ***	0.409856 ***	-0.47378 **	0.670125	-0.07205	0.400598 **
		se	0.331777	0.194396	0.139924	0.118162	0.10714	0.1	0.234346	0.506274	0.263206	0.173218
	RJuice		1.171639 ***	1.410051 ***	1.159159 ***	0.953864 ***	0.649026 ***	0.491572 ***	0.133529	-2.00948 ***	-1.37136 ***	-0.58351 **
		se	0.31616	0.200955	0.167169	0.14802	0.136181	0.131707	0.238351	0.751604	0.374771	0.26011
	RPinneaple		2.264801 ***	2.183513 ***	1.881982 ***	1.672079 ***	1.358864 ***	1.326775 ***	-0.95394 ***	-0.38662	-0.50666	0.122718
		se	0.281705	0.198631	0.165899	0.148439	0.137646	0.131668	0.305366	0.708827	0.384386	0.275159
Information Treatment												
	IRtxred		-0.19085	-0.54223	-0.3489	-0.25686	-0.19676	-0.2294 *	0.248057	-0.26119	-0.21186	-0.17729
		se	0.480884	0.337971	0.224931	0.172646	0.13249	0.118693	0.212326	0.284141	0.17367	0.145619
	IRtxsal		-0.70222 **	-0.31707	-0.17463	-0.06726	-0.07109	-0.00288	-0.28334	0.333384	0.087477	0.102837
		se	0.345352	0.194295	0.130043	0.107651	0.093893	0.086018	0.195285	0.233197	0.14841	0.117302
	IRtrte		0.727744 **	0.645455 ***	0.855963 ***	0.749311 ***	0.495006 ***	0.398153 ***	-0.51575 **	0.027749	-0.3925 *	0.16638
		se	0.318387	0.19014	0.131649	0.108461	0.096154	0.089671	0.203164	0.362501	0.220477	0.142527
	IRtjuice		1.812941 ***	1.858747 ***	1.533919 ***	1.404708 ***	1.044761 ***	0.9219 ***	-1.02502 ***	0.253516	-0.76096 **	0.21075
		se	0.293976	0.186128	0.152623	0.133238	0.122779	0.117376	0.285239	0.473395	0.34742	0.221903
	IRtpinnea		2.457474 ***	2.298671 ***	2.084327 ***	1.89769 ***	1.540859 ***	1.440235 ***	-1.81637 ***	0.757166	-0.7092	0.011809
		se	0.281747	0.18603	0.151167	0.13531	0.125466	0.120923	0.353012	0.587515	0.43886	0.298641
	NOBS [(b)]		2022	3720	5107	6223	7056	7643	2022	587	1420	2536
	Tjur's R sq [(c)]		0.238977	0.168373	0.095758	0.075108	0.048943	0.043887	0.075911	0.133947	0.066051	0.043702
	Pseudo R sq		0.269726	0.197427	0.116509	0.091491	0.05777	0.047842	0.06026	0.075364	0.026351	0.012654
	Chi Squared Test [(d)]		157.27	281.48	279.11	289.25	220.3	203.95	44.27	6.93	9.64	6.89
	P-Value		0	0	0	0	0	0	0	0.2258	0.086	0.2287

(a) Single (*), double (**), and triple (***) asterisks are used to denote significance at the 0.10, 0.50, and 0.01, respectively. (Std Errors)

(b) number of observations of the exploded data

(c) Tjur's coefficient of descrimination to measure goodness of fit (predictive power) of model. It is independent of number of observations

(d) Chi Squared test for structural stability of parameters. The bids were taken as a treatment and their joint significance was tested.

Table 5. Heteroscedastic Conditional Logit Parameter Estimates of the Exploded Data

	Variable		Trank1 [a]	Trank2	Trank3	Trank4	Trank5	Trank6	Bottom1	Brank2	Brank3	Brank4
Variety												
	Texas Red		-0.43955	-0.35614	-0.11613	-0.12271	-0.03361	-0.00546	0.370905	0.01808	-0.12071	-2.2E-05
		se	0.433793	0.341961	0.287206	0.288335	0.255813	0.290422	0.226408	0.04918	0.147654	0.000236
	Texas Salavatski		-0.57456 **	-0.27224	-0.05521	0.08856	0.130303	0.157845	-0.39844 *	0.029887	-0.03593	-8.00E-06
		se	0.285928	0.202373	0.172622	0.174889	0.175095	0.209268	0.21014	0.084755	0.12158	8.65E-05
Product Form												
	Ready to Eat		0.556579 *	0.754344 ***	0.85368 ***	0.762573 ***	0.57741 ***	0.76647 ***	-0.41482 *	0.016313	-0.25994	7.94E-06
		se	0.293629	0.220738	0.195367	0.1913	0.199327	0.237369	0.217271	0.060876	0.219302	8.64E-05
	Juice		1.470412 ***	1.82441 ***	1.656432 ***	1.678722 ***	1.454508 ***	1.616442 ***	-0.57995 *	-0.18713	-1.13275	-6.7E-05
		se	0.313035	0.245125	0.232539	0.2299	0.236817	0.282906	0.314922	0.455452	0.451907 **	0.000717
	Pinneaple		2.323489 ***	2.362831 ***	2.182196 ***	2.141209 ***	1.710128 ***	1.810912 ***	-1.58223 ***	-0.08251	-0.69599	-3.1E-05
		se	0.31208	0.24503	0.232964	0.235295	0.245431	0.300761	0.37238	0.196479	0.346732 **	0.000327
	Scale Parameter		-0.01588	-0.30346 **	-0.62462 ***	-0.86165 ***	-0.54441	-11.7602	-0.5588	2.751475	0.454996	9.895316
		se	0.143753	0.142381	0.21316	0.299236	0.395094	545.4773	0.410304	2.541242	0.497061	10.7304
	NOBS [b]		1592	2379	2462	2214	1676	975	1592	479	1009	1435

(a) Single (*), double (**), and triple (***) asterisks are used to denote significance at the 0.10, 0.50, and 0.01, respectively. (Std Errors)

(b) number of observations of the exploded data