Agricultural & Applied Economics Association

ARE CHOICE EXPERIMENTS INCENTIVE COMPATIBLE? A TEST WITH QUALITY DIFFERENTIATED BEEF STEAKS

JAYSON L. LUSK AND TED C. SCHROEDER

This study compares hypothetical and nonhypothetical responses to choice experiment questions. We test for hypothetical bias in a choice experiment involving beef ribeye steaks with differing quality attributes. In general, hypothetical responses predicted higher probabilities of purchasing beef steaks than nonhypothetical responses. Thus, hypothetical choices overestimate total willingness-to-pay for beef steaks. However, marginal willingness-to-pay for a change in steak quality is, in general, not statistically different across hypothetical and actual payment settings.

Key words: CVM, conjoint analysis, heteroskedastic extreme value model, hypothetical bias, multinomial probit, random parameters logit.

Economists have used choice experiments (CE) and choice-based conjoint analysis to estimate the value of nonmarket goods, the trade-off between environmental or food quality attributes, and market share of novel products (e.g., Adamowicz et al.; Jayne et al.; Layton and Brown; Lusk, Roosen, and Fox; Unterschultz et al.). The frequent use of CEs might be explained by a number of factors. First, CEs are flexible because numerous attributes can be simultaneously valued. Conventional contingent valuation (CV) methods require cumbersome repeated questions to achieve the same level of complexity. Second, CEs are consistent with random utility theory (Ben-Akiva and Lerman) and Lancaster's theory of consumer demand, which posits that consumers derive utility from consumption of attributes embodied in a good. Third, individual CE questions are typically framed in a manner that closely resembles consumer purchasing decisions (e.g., making a choice among alternative products).

Because CE questions closely mirror actual consumer purchasing situations, it has been hypothesized that CEs are less prone to one of the primary drawbacks of the CV method: hypothetical bias in WTP estimates. Individuals often overstate their willingness-to-pay in hypothetical CV situations (Cummings, Harrison, and Rutström; Fox et al.; List and Gallet; List and Shogren). This bias has generated skepticism of the CV method in general (e.g., Diamond and Hausman) and has raised questions about the validity of specific benefit measures used in public policy. Although evidence of bias in hypothetical CV applications is widespread, recent work has suggested that CE responses are statistically indistinguishable across hypothetical and nonhypothetical treatments (Carlsson and Martinsson).[1] Further, many CE applications in the marketing literature implicitly assume preferences are accurately revealed (e.g., Louviere and Woodworth). Aside from Carlsson and Martinsson, little work has examined the external validity of CEs.[2]

_____

[1] List also found no evidence of hypothetical bias in CE responses for some, but not all applications.

[2] Although not a test of external validity or incentive compatibility, a few studies have compared hypothetical CE estimates with other hypothetical conjoint or hypothetical CV estimates (e.g., Adamowicz et al., Boyle et al., Mackenzie). Several other studies have also compared stated CEs preferences with revealed preferences; however, the focus of these analyses was primarily to create a better combined model rather than testing for incentive compatibility (e.g., Adamowicz et al., Hensher, Louviere, and Swait; Louviere, Hensher, and Swait). A few other studies have tested for internal validity of CEs (e.g., Johnson and Mathews).

However, testing whether CEs are subject to hypothetical bias is necessary to determine the reliability and validity of this valuation method. Rising popularity of the CE method elevates the need for more research to assess the merits of this particular valuation approach. Several studies have expressed concerns with CEs. Because subjects are often asked to respond to several repeated CE questions, subject fatigue or learning may affect behavior (Bradley and Daly; Johnson and Desvousges). In addition, DeShazo and Fermo and Swait and Adamowicz found that complexity of the CE influenced consistency of choices and aggregate preferences. These problems might serve to accentuate problems with hypothetical bias: subjects might behave inconsistently when they do not have to back up their choices with real commitments.

This study compares hypothetical and non-hypothetical responses to CE questions. Although many applications use CEs to value public goods, in this application, we focus on the incentive compatibility properties of a CE for private goods: beef ribeye steaks with differing quality attributes. CEs were originally developed in the marketing and transportation literatures and have been used for more than a decade to value private goods. Thus, it is imperative to test for hypothetical bias in CEs with private goods.[3]

Our study generates a number of important findings. In contrast to Carlsson and Martinsson, our estimates suggest that hypothetical CE responses are statistically different from actual CE responses. Responses to the hypothetical CE generally predict higher probabilities of purchasing beef steaks than when payment is actually required. Despite the finding of hypothetical bias, we find that marginal WTP for steak attributes is similar in the hypothetical and nonhypothetical settings.

Experimental Methods

To test the incentive compatibility properties of CEs, a random sample of consumers was recruited to evaluate quality-differentiated beef steaks. Beef has traditionally been marketed as a homogeneous commodity, but some

brands are beginning to appear in the market place (i.e., Laura's Lean, Coleman's Natural, Certified Angus Beef, etc.); however, these brands are not available in all markets and are rarely marketed together in the same store such that relative demand for each of the steaks can be directly analyzed. As such, more research is needed to identify consumer preferences for different steak attributes (see Lusk for a discussion on this issue). Three-quarter pound ribeye steaks with varying quality attributes were chosen as the unit of analysis. Because many of the steak types were either new or unavailable in the local marketplace, demand for the attributes was uncertain. In this regard, this study represents a classical use of CEs to estimate potential market shares of new products and determine relative preferences for product attributes (e.g., Louviere and Woodworth; Louviere 1991).

To test the incentive compatibility of a CE, one must first establish the relevant baseline to compare hypothetical responses against. Following Shogren et al., transaction data from a grocery store might be used in this regard. However, there are problems with this approach. First, based upon our previous experience in working with grocery stores in meat-purchase experiments, finding one that would participate in this particular study would be an arduous task. Second, collecting demographic information from each shopper purchasing steaks in a store setting would pose a major challenge. Without demographic information, one would have to question whether differences in real and hypothetical responses were due to incentive compatibility or due to different sample makeups. Third, the high costs and logistical problems associated with placing new products in a grocery store are prohibitory. Finally, the store environment itself could cause differences in hypothetical and real responses (Lusk and Fox). A comparison of store purchases and responses to a nonhypothetical survey presents a joint test of the hypotheses that the retail environment and actual payment affects valuations. To isolate the potential effect of hypothetical bias, we conducted two treatments in identical environments with the only difference being whether payment was actually required.

A second consideration in designing an experiment to test for incentive compatibility of CEs is whether to conduct within- or between-subject tests. Carlsson and Martinsson used a within-subject design where subjects first responded to a hypothetical CE and then

---

[3] It is important to note that incentives can differ when valuing private as opposed to public goods. Carson, Groves, and Machina contend that conventional CV techniques are not incentive compatible when considering the provision of a private good in a hypothetical context. Thus, our findings regarding the incentive compatibility of CEs may not be extendable to public goods, which is typically the focus of environmental valuation.

Scenario	Steaks					
11		Guaranteed		USDA	Certified	None of
	Generic	Tender	Natural	Choice	Angus Beef	These
	$6.75	$7.88	$9.00	$5.63	$7.88	
	⇩	⇩	⇩	⇩	⇩	⇩
I would choose . . .	☐	☐	☐	☐	☐	☐

Figure 1. Example choice experiment question

responded to a nonhypothetical CE. Although within-subject designs can generate statistical tests with greater power, there are disadvantages to this approach. The primary disadvantage with a within-subject design is that participation in a nonhypothetical treatment can alter subsequent behavior in a hypothetical treatment or vice versa. To mitigate these potential problems, we designed a between-subject experiment where two similar samples of respondents each participated in different treatments: hypothetical or nonhypothetical. In the subsequent analysis, we test for and reject any differences in demographic makeups of our two samples.

Given these considerations, random digit dialing techniques were used to recruit experimental subjects from the general population of a midwestern college town. Subjects were offered $40 cash to participate in a "steak preference experiment," to be conducted in the meat laboratory on the local university campus. Individuals that agreed to participate were assigned a time and date that was convenient for them (two-hour sessions were held several times daily from 27 February 2002 to 3 March 2002). Confirmation letters were mailed to participants one week prior to the sessions to increase attendance rates. Subjects participated in one of two CE treatments: hypothetical or nonhypothetical.

Upon arriving at the session, subjects were paid $40 cash and completed a short demographic questionnaire. Subjects were then shown and provided an opportunity to examine five different beef ribeye steaks: a generic steak, a guaranteed tender steak, a natural steak, a USDA Choice steak, and a Certified Angus Beef® (CAB) steak. An information sheet describing each of the beef steaks was read aloud and distributed to participants.[4] After reading the information sheet and visually examining the steaks, subjects responded to a series of seventeen repeated CE questions.

In each of the CE questions, five ribeye steaks were offered at various price levels. A sample CE question is illustrated in figure 1. Subjects were asked to indicate which steak, or none, they preferred in each scenario. Generic steak prices (12 oz) were varied between $3.38, $4.50, $5.63, and $6.75. Prices of guaranteed tender, natural, Choice, and CAB steaks (12 oz) were varied between $5.63, $6.75, $7.88, and $9.00. Price levels were chosen to encompass the range of steak prices in local grocery stores and retail prices reported by the National Cattlemen's Beef Association and the U.S. Department of Agriculture/Economic Research Service. To be consistent with the vast majority of CE studies, a "none of these" option was also included in each choice set.[5]

In this set-up, there are five steaks (plus none) varied at four different price levels. Subjects would have to be shown $4^5 = 1,024$ different choice sets if presented with every steak at every combination of price levels. To reduce the number of questions respondents had to answer, we generated an orthogonal fractional factorial design. In this design, the prices of each steak are totally uncorrelated with the prices of each of the other four steaks. The resulting design consisted of sixteen choice sets or scenarios. As a basis of comparison, a 17th choice set was added where all steaks were priced at the same level (see the Appendix for the CE scenarios). The set-up of the CE closely resembles that in Erdem and Swait (discussed in more detail in Louviere, Hensher, and Swait).

Subjects assigned to the nonhypothetical treatment were informed that when participants completed all CE questions, one of the seventeen questions would be randomly selected as binding by having one of the participants draw a number out of an envelope. After the binding scenario was determined, each

---

[4] Complete experimental instructions and the information sheet are available from the authors upon request.

[5] A "none of these" option was included to set the origin or base of the utility scale from which other alternatives are compared (Louviere 1988). More pragmatically, inclusion of a "none" option makes the choice task more realistic as consumers are able to choose this option when shopping.

participant purchased the steak they chose in the binding scenario. In the experimental instructions, subjects were explicitly told that actual payment would occur for the binding scenario and that they should evaluate each scenario carefully, as all scenarios had an equally likely chance of being binding. Subjects were also told that marking the "none of these" option was acceptable, and in the case that "none of these" was marked in a binding scenario, no purchase would be made.

For the hypothetical treatment, procedures and CE questions were identical except subjects were told that payment for a steak would not actually take place. Although payment was not required, we asked subjects to behave as if choices were real and binding. Subjects were asked to take time to evaluate each scenario as if they were going to actually have to face the consequences of their choice. Participants were instructed to behave, for example, as if there were an equally likely chance they would be held accountable for their choice in scenario 3 as if they were to be held accountable for their choice in scenario 14. Subjects were told that marking the "none of these" option was acceptable in any scenario. Lastly, subjects in the hypothetical treatment were asked to indicate their preferences under the assumption that only one steak would be purchased.

Analysis

A number of methods could be used to compare responses across hypothetical and non-hypothetical treatments. First, the frequency of steak choices in each scenario can be compared across treatments. However, testing whether choice frequencies are equivalent across treatments requires that each steak be chosen a sufficient number of times in each CE scenario. In other words, if a small percentage of individuals choose a particular steak in a particular scenario, traditional chi-squared tests of independence across treatments are not valid. Nevertheless, summarizing differences in choice frequencies across treatments can be informative.

Econometric Models

Econometric model estimates from each treatment can also be compared to identify whether preferences are similar in hypothetical and nonhypothetical CEs. In each CE question, respondents had to choose between five steaks and the "none of these" option. Let the ith

consumer's utility of choosing option steak j be given by

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

where V_{ij} is the systematic portion of the utility function determined by the steak attributes and ε_{ij} is a stochastic element. Assuming V_{ij} is linear in parameters, the functional form of the utility function for alternative j can be expressed as

$$(2) \qquad V_{ij} = \beta_j + \alpha_j P_{ij}$$

where j = generic, guaranteed tender, natural, Choice, and CAB steaks, P_{ij} is the price of alternative j for consumer i, β_j are coefficients representing alternative specific constants for each of the steak alternatives relative to the "none of these" option, and α_j are coefficients representing the effect of the jth steak price on utility for the jth steak. The probability that a consumer chooses alternative j is

$$\text{Prob}\{V_{ij} + \varepsilon_{ij} \geq V_{ik} + \varepsilon_{ik}; \quad \text{for all } k \in C_i\}$$

where C_i is the choice set for respondent i, that is, $C_i = \{$generic, guaranteed tender, natural, USDA Choice, CAB, none$\}$. In the following subsections, several alternative econometric models are discussed that make different assumptions about the form of equation (2) or the distribution of ε_{ij} in equation (1).

Multinomial Logit Model

If the ε_{ij}'s are independently and identically distributed across the j alternatives and N individuals with an extreme value distribution, Louviere, Hensher, and Swait show that the probability of consumer i choosing alternative j is given by the multinomial logit (MNL) model:

$$(3) \qquad \text{Prob}\{j \text{ is chosen}\} = \frac{e^{\mu V_{ij}}}{\sum_{k \in C} e^{\mu V_{ik}}}$$

where μ is a scale parameter that is inversely related to the variance of the error term.

Although the scale parameter is unidentifiable within any particular data set, one can estimate a relative scale parameter across data sets (Swait and Louviere). Identifying the relative scale parameter is important in determining whether differences in parameter estimates across data sets are due to differences in variance or differences in underlying preferences. In this application, an artificial nested logit

model is used to identify the relative scale parameter between hypothetical and non-hypothetical treatments (Adamowicz et al.; Hensher and Bradley; Hensher, Louviere, and Swait). Preference equality is tested by controlling for differences in scale and estimating a model that imposes the null hypothesis of parameter equality across treatments (see Louviere, Hensher, and Swait, p. 364).

Universal Logit Model

Despite the prevalent use of the MNL in the literature and the relative ease with which welfare estimates can be calculated from it, the MNL is derived based on the independence of irrelevant alternatives (IIA) assumption. The IIA assumption results from the assumption that model errors are independently distributed across alternatives. To test for potential violations of the IIA assumption, we also estimated a universal or "mother" logit model (McFadden, Train, and Tye).[6]

The universal logit model is the most general model in the generalized extreme value family of models, which includes the MNL and nested logit models. In a universal logit model, the utility of each alternative is specified as a function of the attributes of each of the other alternatives. In this application, the utility of the generic steak, for example, is a function of an alternative specific constant and the prices of the generic, guaranteed tender, natural, Choice, and CAB steaks. The functional form of the utility function for alternative j in the universal logit can be expressed as

$$(4) \qquad V_{ij} = \beta_j + \sum_{k=1}^{5} \alpha_{jk} P_{ik}$$

where j = generic, guaranteed tender, natural, Choice, and CAB steaks, k = generic, guaranteed tender, natural, Choice, and CAB steaks, P_{ik} is the kth steak's price for consumer i, and α_{jk} represents the effect of the kth steak's price on the utility for the jth steak. In the universal logit specification, the utility function for each alternative now has own- and cross-price effects, which clearly incorporates any violations of the IIA assumption that

might be present in the MNL model. By testing the joint hypothesis that all cross-price effects are zero, the IIA assumption is directly tested (McFadden, Train, and Tye). To estimate the universal logit, equation (4) is simply substituted, in place of equation (2), into equation (3). The procedure previously discussed is still applicable to identify the relative scale parameter between hypothetical and nonhypothetical treatments with the universal logit.

If violations of the IIA property are found, a number of alternative models can be estimated that do not assume errors are independently distributed across alternatives.[7] In this article, we consider three alternative specifications: the heteroskedastic extreme value (HEV) model, the multinomial probit (MNP) model, and the random parameters logit (RPL) model. Each of these three models relaxes the IIA property of the MNL, and we present results associated with all three models to increase the robustness of our conclusions and to illustrate how welfare measures might differ across alternative specifications.

Heteroskedastic Extreme Value Model

In contrast to the MNL and Universal Logit, which assume that error variance is equivalent across steak alternatives, the HEV model allows error variance to differ across alternatives. In the HEV model, the utility function is the same as that given in equation (1), but the ε_{ij} are assumed independently but not identically distributed for each steak alternative. Given this assumption, Louviere, Hensher, and Swait show the probability of choice in the HEV is

$$(5) \qquad \text{Prob}\{j \text{ is chosen}\}$$

$$= \int_{-\infty}^{\infty} \prod_{k \in C, k \neq j} F\left[\frac{V_j - V_k + \varepsilon_j}{\mu_k}\right]$$

$$\times \frac{1}{\mu_j} f\left(\frac{\varepsilon_j}{\mu_j}\right) d\varepsilon_j$$

where $F(\cdot)$ is the standard cdf of the extreme value distribution, $f(\cdot)$ is the pdf of the extreme value distribution, and μ_j is the scale parameter for alternative j that is inversely related to the standard deviation of the error component

---

[6] Although the issue is not explored in this article, agribusiness managers are likely to be quite interested in substitutability between steak types. The MNL imposes a very strict structure on cross-price elasticities, whereas the universal logit permits a very flexible structure on cross-price elasticities (see Krishnamurthi, Raj, and Sivakumar for managerial implications derived from the universal logit).

[7] Although the universal logit relaxes the IIA assumption, it may be inconsistent with utility maximization. McFadden outlines conditions an estimated model must meet to be consistent with random utility maximization; however, it is difficult to determine *ex post* whether an estimated model satisfies such conditions.

of alternative j. Methods for computing the integral in equation (5) are outlined in Bhat and Louviere, Hensher, and Swait. The HEV model permits estimation of a scale parameter for each choice alternative, with the scale of one alternative set equal to one for identification purposes. Because different alternatives have different scales, the IIA assumption is relaxed. Applications employing the HEV can be found in Allenby and Ginter; Bhat; Louviere, Hensher, and Swait. As shown in Hensher, Louviere, and Swait, the HEV can be used to combine data sets (such as the case with our hypothetical and nonhypothetical data sets) by controlling for differences in variance across treatment at the alternative level.

Multinomial Probit

The MNP model also permits error variance to differ across alternatives, but instead assumes errors given in equation (1) are normally distributed. In theory the MNP model is more general than the HEV model because the MNP model permits estimation of the entire error covariance matrix instead of just alternative-specific variances, as is the case with the HEV model. As shown by Louviere, Hensher, and Swait, the probability of choice in the MNP model is given by

$$
(6) \quad
\begin{aligned}
&\text{Prob}\{j \text{ is chosen}\} \\
&= \int_{-\infty}^{\varepsilon_j + V_j - V_k} \cdots \int_{-\infty}^{\infty} \cdots \\
&\quad \int_{-\infty}^{\varepsilon_j + V_j - V_J} f(\varepsilon)\, d\varepsilon_J \cdots d\varepsilon_1
\end{aligned}
$$

where J is the choice set size and $f(\cdot)$ is a J-variate normal density function with mean zero and $J \times J$ covariance matrix Σ. Computing the probabilities in equation (5) can be computationally intensive because of the need to evaluate multiple integrals and identification of all parameters in Σ can be a challenge. To put the MNP into practice, we assume all off-diagonal correlations in Σ are zero. To further ensure identification, standard deviations of two alternative-specific errors are restricted to one (see discussion in Greene). This produces a model very similar to the HEV model, but with errors assumed normally distributed instead of extreme value. Because error variances are estimated for each alternative (up to identification), the IIA assumption is relaxed. The MNP model was originally discussed in

Hausman and Wise and applications can be found in Louviere, Hensher, and Swait. As was the case with the HEV model, the MNP model can be used to combine data sets by controlling for difference in variance at the alternative level.

Random Parameters Logit

The last model we consider is the RPL. The RPL is based on a modification of the MNL, but instead of taste parameters assumed constant in the population, as is the case in the MNL, the RPL allows taste parameters to vary in the population. The probability of choice in the RPL model is given by

$$
(7) \quad \text{Prob}\{j \text{ is chosen} \mid \lambda_i\} = \frac{e^{V_{ij}}}{\sum_{k \in C} e^{V_{ik}}}
$$

where λ_i is an individual-specific random disturbance of unobserved heterogeneity. In general, the coefficient vector for individual i in the RPL is $\beta_i = \bar{\beta} + \sigma \lambda_i$, where $\bar{\beta}$ is the population mean, σ is the standard deviation of the marginal distribution of β, and λ_i is a random term assumed normally distributed mean zero and unit standard deviation. If $\sigma = 0$, then the RPL is equivalent to the MNL. Aside from relaxing the IIA assumption, the RPL is advantageous because random-effects introduced by the panel-nature of the data (i.e., each individual responded the seventeen choice questions) can be incorporated quite easily. Further computational details on the RPL model can be found in Greene; Layton and Brown; Louviere, Hensher, and Swait; Train; and Revelt and Train.

In our application of the RPL, we assume that the alternative-specific constants are independently normally distributed in the population. Following Layton and Brown and Revelt and Train, price coefficients were assumed fixed in the population.[8] In theory, it is possible to combine both hypothetical and nonhypothetical responses and estimate the relative scale-parameter between the two data sources with the RPL as was the case with the MNL. However, the likelihood function is not globally concave over the relative scale parameter with the RPL, and in practice we found the likelihood function to be quite erratic over

---

[8] Fixing the price coefficients ensures that estimated willingness-to-pay will be normally distributed. It also forces all respondents to have a negative price coefficient, which would not be possible if the coefficient were assumed normally distributed. Further, we found that convergence was a problem if all variables were allowed to be random.

relevant ranges of relative scale. Thus, in the analysis that follows, we do not control for differences in scale across hypothetical and non-hypothetical data sets in the RPL estimation. As a result, our test of whether RPL model parameters are equivalent across hypothetical and nonhypothetical data sets is a test of the joint hypothesis of equality of both taste parameters and scale. Despite this weakness, comparisons of hypothetical and nonhypothetical WTP calculated from the RPL are valid because the scale parameter is constant within each data set and is cancelled out in calculation of WTP.[9]

Willingness-to-Pay

WTP estimates are derived by determining the price difference necessary to invoke indifference between two steak alternatives. Total WTP to obtain steak alternative j versus "none of these" is simply calculated as the ratio of the alternative specific constant to the price coefficient: $-\beta_j/\alpha_j$.[10] Marginal WTP for steak j versus steak k can be calculated in two ways. First, WTP to exchange steak alternative j for steak alternative k can be determined by finding the price of k that satisfies the equality: $V_j = V_k$. A problem with this approach is that it can generate asymmetrical valuations, with $WTP_{jk} \neq WTP_{jk}$. Further, this approach generates marginal WTP that is conditional upon the price of other alternatives. To mitigate these problems, we calculate marginal WTP for steak j versus steak k as: $-\beta_j/\alpha_j + \beta_k/\alpha_k$. That is, marginal WTP is determined by subtracting total WTP for steak j from total WTP for steak k.

Standard deviations of WTP are derived by generating a distribution of 1,000 WTP estimates using the parametric bootstrapping method proposed by Krinsky and Robb. More specifically, 1,000 observations were drawn from multivariate normal distributions created using the coefficient estimates the variance-covariance matrices from each of the econometric models. To test whether there are differences in hypothetical and nonhypothetical willingness-to-pay, we used the combinatorial test mentioned in Poe, Severance-Lossin, and Welsh and investigated further in Poe, Giraud, and Loomis. The combinatorial approach provides a simple unbiased, nonparametric test of the difference in two distributions. The combinatorial test is carried out by comparing the 1,000 bootstrapped values from hypothetical WTP for a given steak and econometric model to the 1,000 bootstrapped values for the corresponding nonhypothetical WTP for the same steak and econometric model. More specifically, the difference between hypothetical WTP and nonhypothetical WTP is calculated for all possible combinations of the bootstrapped values—that is, we calculate ((hypothetical WTP)$_t$ − (nonhypothetical WTP)$_z$) for all t and z, where $t = 1,000$ bootstrapped hypothetical WTP values and $z = 1,000$ nonhypothetical bootstrapped WTP. As such, 1,000,000 differences are calculated for each hypothesis test of interest. The proportion of differences that are less than zero is the p-value associated with the one-sided test that the hypothetical WTP distribution is greater than the nonhypothetical WTP distribution.[11]

Tests for WARP Violations

As a final comparison of hypothetical and nonhypothetical responses, the internal consistency of an individual's responses across the seventeen CE questions can be tested. Because previous studies have observed inconsistencies across CE questions (e.g., DeShazo and Fermo), it is important to test whether such an effect is present here. To test for consistency across CE questions, violations of the weak axiom of revealed preferences (WARP) were checked using a nonparametric approach. Consistency with WARP is a necessary condition that an individual's preferences are stable over the seventeen CE responses. In its simplest form, a WARP violation occurs if an individual chooses a particular steak (e.g., steak A) in a scenario when another steak (e.g., steak B) is less expensive, but chooses steak B in a different scenario when steak A was less expensive. Procedures outlined in Chalfant and Alston were employed to test for violations of WARP (see Varian 1982, 1983 for further discussion on the method). Of interest here is whether a greater frequency of individuals violate WARP when they are not held accountable for their choices as when real payment occurs.

---

[9] All models reported in this article were estimated using LIMDEP.

[10] This WTP measure assumes no uncertainty regarding consumer choice; see Adamowicz et al. for an alternative welfare measure that incorporates choice uncertainty.

[11] Poe, Giraud, and Loomis provide simple computer code to carry out the test.

Table 1. Summary Statistics of Selected Participant Demographics by Treatment

Variable	Definition	Treatment[a]	
		Nonhypothetical	Hypothetical
Gender	1 = female; 0 = male	0.42[b]	0.41
		(0.50)[c]	(0.50)
Age	Age in years	33.13	29.62
		(17.55)	(14.92)
Education	1 = high school; 2 = some college; 3 = technical school;	3.66	3.38
	4 = associate's degree; 5 = bachelor's degree;	(2.06)	(1.98)
	6 = master's degree; 7 = J.D.; 8 = doctorate		
Student	1 = full time student; 0 otherwise	0.51	0.60
		(0.50)	(0.50)
Income	Household income level	4.63	3.76
	1 = less than $10,000; 2 = $10,000 to 19,999...	(4.82)	(4.07)
	19 = $180,000 to $189,999; 20 = more than $190,000		
Number of participants		67	37

[a] The null hypothesis of equality of means across treatments cannot be rejected for any of the demographics.
[b] Reported statistics are mean values.
[c] The numbers in parentheses are standard deviations.

Results

Subjects were recruited as part of a larger project on consumer demand for quality differentiated beef. Response rates for the overall project were as follows: 45% of random digit dials were working numbers, 48% of contacted individuals agreed to participate in a research session, and 85% of individuals who agreed to take part in a session actually participated. Overall, sixty-seven subjects participated in a nonhypothetical CE treatment and thirty-seven individuals participated in a hypothetical CE treatment.[12] Table 1 reports summary statistics of basic demographics for the two treatments. For each of the demographics reported in table 1, the null hypothesis of equality of means across treatments cannot be rejected.

Summary statistics of responses to each of the seventeen CE questions are reported in table 2 (prices for the seventeen choice sets are listed in the Appendix). One result is readily apparent: participants in the nonhypothetical treatment chose the "none" option much more frequently than participants in the hypothetical treatment. For all seventeen scenarios, the percentage of individuals choosing "none" was greater in the nonhypothetical treatment, with

the smallest difference between actual and hypothetical "none" frequencies being 8.31% (scenario 4) and the largest difference being 32.84% (scenario 7). As expected, many of the scenarios contain steak options that were never or infrequently chosen, precluding the use of chi-squared tests of independence to test for differences in choice frequencies across treatment.

Table 3 reports MNL estimates for pooled and segmented samples. To preserve orthogonality of prices across steak options, only the first sixteen CE scenarios shown in the appendix were included in the estimation. The joint model restricts steak preference parameters to be equal across treatment and allows for estimation of the relative scale parameter. Results indicate that the scales of hypothetical and nonhypothetical data are statistically equivalent (i.e., the relative scale is not statistically different from one). The null hypothesis of preference equality across treatments is $\beta_{nonhypothetical} = \beta_{hypothetical}$. As shown in Louviere, Hensher, and Swait, the test for preference regularity is $-2(LL_J - \sum LL_i)$, which is distributed χ^2 with $K(M-1)$ degrees of freedom, where LL_J is the log likelihood value for the pooled model after controlling for scale, LL_i are the log likelihood values of separate MNL models from each treatment, K is the number of restrictions, 10, and M is the number of treatments, 2. For these data, parameter equality is strongly rejected ($\chi^2 = 106.8$; $p < 0.01$). Thus, we reject the hypothesis of equality of hypothetical and nonhypothetical CE responses.

---

[12] There were three sessions of the nonhypothetical treatment on Thursday at 5:10 pm (thirty-five participants), Friday at 2:00 pm (twenty-two participants), and Sunday at 5:10 pm (ten participants). There were three sessions of the hypothetical treatment on Wednesday at 8:00 pm (sixteen participants), Sunday at 2:20 pm (fifteen participants), and Sunday at 6:30 pm (six participants).

Table 2. Comparison of Hypothetical and Nonhypothetical Choices in Seventeen Shopping Scenarios

Treatment	Percent Choosing	Scenario							
		1	2	3	4	5	6	7	8
Nonhypothetical	Generic steak	28.4	10.4	0.0	0.0	0.0	10.4	16.4	0.0
	Guaranteed tender steak	1.5	1.5	0.0	1.5	3.0	1.5	9.0	61.2
	Natural steak	4.5	1.5	29.9	0.0	3.0	1.5	16.4	6.0
	USDA choice steak	19.4	65.7	0.0	7.5	1.5	1.5	25.4	0.0
	Certified Angus Beef steak	20.9	1.5	32.8	74.6	76.1	68.7	0.0	3.0
	None	25.4	19.4	37.3	16.4	16.4	16.4	32.8	29.9
Hypothetical	Generic steak	48.6	10.8	0.0	2.7	0.0	21.6	24.3	0.0
	Guaranteed tender steak	0.0	0.0	2.7	0.0	13.5	0.0	13.5	64.9
	Natural steak	10.8	2.7	48.6	0.0	10.8	0.0	40.5	16.2
	USDA choice steak	13.5	86.5	0.0	5.4	0.0	0.0	21.6	2.7
	Certified Angus Beef steak	24.3	0.0	37.8	83.8	67.6	75.7	0.0	2.7
	None	2.7	0.0	10.8	8.1	8.1	2.7	0.0	13.5

Treatment	Percent Choosing	Scenario								
		9	10	11	12	13	14	15	16	17
Nonhypothetical	Generic steak	11.9	0.0	0.0	11.9	40.3	10.4	34.3	0.0	0.0
	Guaranteed tender steak	3.0	0.0	1.5	38.8	0.0	13.4	31.3	50.7	7.5
	Natural steak	25.4	1.5	0.0	0.0	1.5	6.0	0.0	3.0	6.0
	USDA choice steak	16.4	73.1	79.1	1.5	10.4	13.4	0.0	16.4	19.4
	Certified Angus Beef steak	1.5	7.5	1.5	17.9	14.9	44.8	3.0	0.0	52.2
	None	41.8	17.9	17.9	29.9	32.8	11.9	31.3	29.9	14.9
Hypothetical	Generic steak	16.2	2.7	0.0	13.5	56.8	32.4	48.6	0.0	0.0
	Guaranteed tender steak	5.4	0.0	0.0	59.5	0.0	10.8	32.4	59.5	18.9
	Natural steak	51.4	0.0	0.0	2.7	0.0	13.5	0.0	0.0	13.5
	USDA choice steak	13.5	89.2	91.9	0.0	5.4	21.6	0.0	27.0	32.4
	Certified Angus Beef steak	0.0	0.0	0.0	16.2	21.6	21.6	5.4	0.0	32.4
	None	13.5	8.1	8.1	8.1	16.2	0.0	13.5	13.5	2.7

Table 3. Multinomial Logit Model Estimates: Comparison of Hypothetical and Nonhypothetical Treatments

Independent Variable	Model		
	Joint	Nonhypothetical[a]	Hypothetical[a]
Alternative specific constants			
Generic (GEN)	4.27*	3.79*	5.28*
	(0.45)[b]	(0.55)	(0.62)
Guaranteed tender (GT)	9.29*	9.28*	9.56*
	(0.80)	(0.93)	(1.10)
Natural (NAT)	7.39*	6.46*	8.65*
	(0.93)	(1.22)	(1.26)
USDA Choice (CHO)	10.81*	9.90*	12.77*
	(0.78)	(0.76)	(1.12)
Certified Angus Beef (CAB)	10.69*	11.48*	9.75*
	(0.73)	(0.81)	(0.94)
Own-price effects			
GEN price	−1.06*	−1.09*	−1.03*
	(0.11)	(0.14)	(0.14)
GT price	−1.46*	−1.52*	−1.35*
	(0.13)	(0.15)	(0.18)
NAT price	−1.28*	−1.24*	−1.28*
	(0.15)	(0.20)	(0.21)
CHO price	−1.59*	−1.51*	−1.75*
	(0.12)	(0.12)	(0.17)
CAB price	−1.56*	−1.72*	1.30*
	(0.11)	(0.13)	(0.14)
Scale parameter[c]	0.90*		
	(0.06)		
Log Likelihood	−1942.2	−1261.4	−627.4
Number of observations[d]	1664	1072	592

*Statistical significance at the 0.05 level or lower.
[a] Estimates are adjusted for differences in scale.
[b] Numbers in parentheses are standard errors.
[c] Scale of nonhypothetical responses is relative to scale of hypothetical responses (restricted to equal one).
[d] Each respondent answered seventeen CE questions, sixteen of which were used in the estimation.

Because of potential violations of the IIA assumption, we also estimated universal logit models.[13] The IIA assumption is tested by comparing the likelihood function values from the MNL models in table 3 to those from the universal logit. The likelihood function values for the universal logit were as follows: joint model = −1908.8; nonhypothetical model = −1245.8; and hypothetical model = −597.7. Likelihood ratio tests indicate that the IIA assumption is rejected for the joint model ($\chi^2 = 66.9$; $p < 0.01$), the nonhypothetical model ($\chi^2 = 31.3$; $p = 0.051$), and the hypothetical model ($\chi^2 = 59.5$; $p < 0.01$). Because of the generality of the universal logit, we also tested for equality of taste parameters across nonhypothetical and hypothetical treatments with this specification. Employing the

test for preference regularity outlined above indicates that nonhypothetical responses are statistically different from hypothetical responses ($\chi^2 = 130.7$; $p < 0.01$). Thus, even after controlling for differences in variance and allowing for a more general utility specification, we reject the hypothesis that subject behavior was equivalent in hypothetical and nonhypothetical treatments.[14]

Although the universal logit is a very general model of consumer behavior, it can be inconsistent with random utility maximization. As such, we report results for the HEV, MNP, and RPL models as shown in table 4. For each estimation method, we report three models: a joint

[14] Following McFadden, Train, and Tye, we introduced the demographic variables reported in table 1 into the universal logit to attempt to mitigate the IIA violation. Even when demographic variables were introduced into the model, the IIA assumption is still violated and we still reject the hypothesis of equality of hypothetical and nonhypothetical responses.

[13] To conserve space, estimates from the universal logit models, which contain thirty coefficients each, are not reported in this article. These results are available from the authors upon request.

Table 4. Comparison of Hypothetical and Nonhypothetical Data: HEV, MNP, and RPL Models

Independent Variables	Heteroscedastic Extreme Value Models			Multinomial Probit Models			Random Parmaters Logit Models		
	Joint	Non-hyp	Hyp	Joint	Non-hyp	Hyp	Joint	Non-hyp	Hyp
Alternative specific constants									
Generic	3.90*	3.79*	12.33*	5.68*	2.23*	8.53*	5.54*	4.01*	11.30*
	(0.33)	(1.31)	(4.73)	(0.92)	(0.47)	(2.68)	(0.78)	(0.73)	(3.03)
GT	4.20*	4.27*	6.79*	7.44*	4.04*	7.13*	8.97*	8.49*	11.21*
	(0.37)	(0.80)	(1.52)	(1.02)	(0.70)	(1.40)	(0.73)	(0.86)	(1.79)
Natural	3.97*	3.54*	6.57*	6.22*	3.91*	10.96*	7.81*	5.96*	12.57*
	(0.34)	(1.09)	(1.56)	(0.94)	(1.40)	(3.99)	(1.26)	(1.18)	(2.83)
Choice	4.582*	5.69*	7.95*	7.84*	3.96*	8.48*	10.38*	9.09*	14.85*
	(0.54)	(1.05)	(1.63)	(0.91)	(0.52)	(1.59)	(0.61)	(0.72)	(1.43)
CAB	4.81*	7.55*	7.17*	8.66*	6.57*	8.00*	10.40*	10.94*	11.68*
	(0.62)	(1.52)	(1.40)	(0.94)	(0.41)	(0.89)	(0.60)	(1.10)	(1.29)
Own-price effects									
Generic	−0.33*	−1.10*	−3.27*	−1.41*	−0.63*	−2.13*	−1.51*	−1.19*	−2.76*
	(0.14)	(0.46)	(1.46)	(0.27)	(0.15)	(0.77)	(0.23)	(0.22)	(0.87)
GT	−0.22*	−0.61*	−0.85*	−1.06*	−0.63*	−1.00*	−1.41*	−1.39*	−1.61*
	(0.08)	(0.15)	(0.27)	(0.17)	(0.12)	(0.23)	(0.12)	(0.14)	(0.29)
Natural	−0.22*	−0.64*	−0.87*	−1.00*	−0.80*	−1.79*	−1.40*	−1.16*	−1.96*
	(0.09)	(0.23)	(0.30)	(0.17)	(0.29)	(0.72)	(0.24)	(0.21)	(0.49)
Choice	−0.27*	−0.80*	−0.91*	−1.03*	−0.55*	−1.12*	−1.52*	−1.38*	−2.03*
	(0.11)	(0.18)	(0.26)	(0.14)	(0.09)	(0.24)	(0.09)	(0.11)	(0.21)
CAB	−0.30*	−1.11*	−0.80*	−1.16*	−0.98*	−1.07*	−1.52*	−1.65*	−1.56*
	(0.12)	(0.26)	(0.22)	(0.11)	(0.06)	(0.13)	(0.09)	(0.18)	(0.19)
Scale/standard deviation parameters[a]									
Non-hyp generic	3.36*	0.94*		2.01*	1.01*		1.73*	1.09*	
	(1.41)	(0.34)		(0.37)	(0.22)		(0.42)	(0.50)	
Non-hyp GT	5.56*	2.15*		1.13*	0.67*		0.18	0.06	
	(2.07)	(0.64)		(0.20)	(0.14)		(0.49)	(0.51)	
Non-hyp natural	5.11*	1.41*		1.30*	1.36*		1.15*	0.67	
	(2.14)	(0.43)		(0.23)	(0.43)		(0.55)	(0.65)	
Non-hyp choice	5.09*	1.68*		0.84*	0.26		0.06	0.16	
	(2.14)	(0.42)		(0.17)	(0.19)		(0.22)	(0.30)	
Non-hyp CAB	4.109*	1.16*		1.09*	1.00		0.03	0.61	
	(1.66)	(0.30)		(0.14)			(0.33)	(0.48)	
Non-hyp None	0.49*	1.00		2.65*	1.00		–	–	
	(0.03)			(0.26)					
Hyp generic	2.70*		0.31*	2.58*		3.04*	1.73*		3.67*
	(1.12)		(0.12)	(0.48)		(1.05)	(0.42)		(1.29)
Hyp GT	5.97*		1.36*	1.08*		1.03*	0.18		0.84
	(2.31)		(0.46)	(0.22)		(0.28)	(0.49)		(0.58)
Hyp natural	4.06*		1.15*	1.74*		2.20*	1.15*		1.78*
	(1.72)		(0.41)	(0.31)		(0.81)	(0.55)		(0.73)
Hyp choice	6.27*		2.57*	0.41*		0.62*	0.06		0.23
	(2.88)		(1.00)	(0.17)		(0.27)	(0.22)		(0.42)
Hyp CAB	5.45*		2.66*	1.00		1.00	0.03		0.46
	(2.42)		(1.14)				(0.33)		(0.39)
Hyp none	1.00		1.00	1.00		1.00	–		–
Log Likelihood	−2031	−1272	−612	−1899	−1257	−614	−1937	−1259	−615
# Obs	1664	1072	592	1664	1072	592	1164	1072	592
Test for equality	294.8*				55.2*				125.4*

*Statistical significance at the 0.05 level or lower.

[a] Estimates are scale parameters of alternative specific errors in the heteroscedastic extreme value models, standard deviations of alternative specific errors in the multinomial probit models, and standard deviations of the alternative specific constants in random parameters logit models, respectively.

model combining hypothetical and nonhypothetical responses, a model with nonhypothetical data only, and a model with hypothetical data only. The first three columns of results in table 4 pertain to the HEV model. The joint HEV model restricts alternative-specific and own-price effects to be equal across treatments, but allows the scale (error variance) to differ across alternative and treatment. Because the scale is inversely related to error variance, a higher scale implies a greater error variance for the alternative. An examination of the scale estimates indicates more noise exists in the generic steak and none alternatives for both the hypothetical and nonhypothetical treatments than for any of the other steak alternatives. The ratio of nonhypothetical and hypothetical scale parameters is 1.24, 0.93, 1.26, 0.81, 0.75, and 0.47 for the generic, guaranteed tender, natural, Choice, CAB, and "none" alternatives, respectively, implying a greater error variance for some alternatives in the hypothetical setting and a greater error variance for some alternatives in the nonhypothetical setting. The most noticeable difference is with respect to the "none" option in the nonhypothetical treatment, which exhibited greater noise relative to other steak alternatives in either treatment. The next two columns report separate HEV models for the nonhypothetical and hypothetical data, respectively. A likelihood ratio test strongly rejects the hypothesis that hypothetical and nonhypothetical taste parameters are equal even after accounting for differences in variance at the alternative level.

The next set of estimates corresponds to the MNP model. Again, we first report a joint model that imposes equality of taste parameters across treatment, but allows differences in variance across alternative. Rather than estimating scale parameters, the MNP directly estimates the standard deviation of the alternative-specific error component. For identification purposes, two standard deviations must be set equal to one. Results from the MNP are generally consistent with the HEV model; the generic and "none" alternatives are associated with the greatest noise. Again, a likelihood ratio test rejects the hypothesis that hypothetical and nonhypothetical taste parameters are equal even after accounting for differences in variance at the alternative level.

The last three columns in table 4 report results from the RPL specification. Results suggest that consumers exhibited significant heterogeneity in preferences for generic and natural steaks, but were relatively homogeneous in their preferences for the other steak

types. Because of the difficulties previously mentioned, a relative scale parameter was not estimated to determine whether there were significant differences in variance across hypothetical and nonhypothetical responses. Nonetheless, a likelihood ratio test rejects the joint hypothesis of equivalence of hypothetical and nonhypothetical taste parameters and scale.

The estimation results from all models overwhelming indicate that hypothetical and nonhypothetical CE generated different taste parameters. This result is robust to alternative specifications of the utility function as shown by the universal logit, and is robust to alternative specifications of unobserved heterogeneity as shown by the HEV, MNP, and RPL models. To illustrate the differences in hypothetical and nonhypothetical responses, we also plotted predicted probabilities of purchase or "market shares" for the generic steak given the HEV model estimates. Figure 2 illustrates nonhypothetical and hypothetical simulated market shares for the generic steak in the hypothetical and nonhypothetical contexts. The hypothetical market share is clearly greater than nonhypothetical market shares at all price levels for the generic steak. Similar figures from other steaks and econometric models are available from the authors.

As another means of evaluating the incentive compatibility properties of the CE, WTP estimates were compared across hypothetical and nonhypothetical treatments for each of the models estimated (table 5). Results indicate that total WTP to obtain each of the five ribeye steaks versus having no steak at all is greater in the hypothetical than in the nonhypothetical setting. This difference is statistically significant for all steaks except the generic steak in the HEV model according to the nonparametric combinatorial test. The direction of bias is consistent with the extant literature on hypothetical bias. List and Gallet, in a summary of twenty-nine experimental studies, reported that average subjects overstate their willingness-to-pay in hypothetical settings by a factor of 3. Data in table 5 indicate that subjects overstated their WTP, on average across steaks, by a factor of about 1.2, regardless of estimation method. Although hypothetical bias exists in this CE, it appears small relative to the average of other studies.

Despite the fact that hypothetical WTP is greater than actual WTP, marginal WTP to exchange steaks with different attributes is, in most cases, similar across treatment. The most notable exception is marginal WTP to
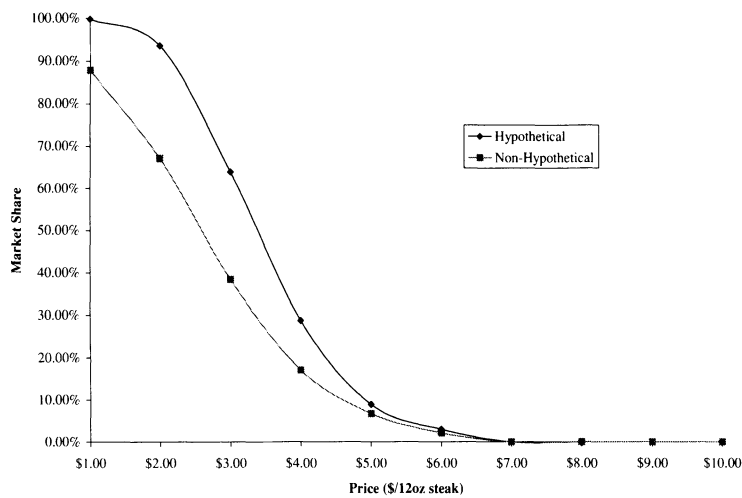
Figure 2. Predicted market shares for generic steak: HEV model

exchange a generic steak for the other steaks as calculated from the HEV model. Although the nonparametric combinatorial test indicates these differences are statistically significant for the HEV model, this result is not robust across the MNP and RPL models. For most of the cases reported in table 5, marginal WTP in the hypothetical context was not statistically different than marginal willingness-to-pay when actual payment was required. Thus, although overall WTP for steaks is greater in

the hypothetical treatment than in the nonhypothetical treatment, marginal WTP between steaks in the hypothetical setting is similar to that in the nonhypothetical treatment.

The last comparison of hypothetical and nonhypothetical responses involved testing for WARP violations. The frequency of WARP violations was extremely low for both treatments. Only four individuals (5.8%) in the nonhypothetical treatment and two individuals (5.4%) in the hypothetical treatment

Table 5. Comparison of Nonhypothetical and Hypothetical Willingness-to-Pay

Willingness-to-Pay for...[a]	Heteroscedastic Extreme Value Models			Multinomial Probit Models			Random Parameters Logit Models		
	Non-hyp[b]	Hyp[b]	p-Value[c]	Non-hyp[b]	Hyp[b]	p-Value[c]	Non-hyp[b]	Hyp[b]	p-Value[c]
Total willingness-to-pay									
Generic versus none	$3.46	$3.77	0.27	$3.39	$4.09	0.09	$3.54	$4.01	0.01
GT versus none	$6.98	$7.99	0.11	$6.12	$6.97	0.02	$6.44	$7.11	0.00
Natural versus none	$5.54	$7.52	0.03	$5.13	$6.43	0.02	$4.90	$6.13	0.00
Choice versus none	$7.09	$8.74	0.01	$6.59	$7.31	0.10	$7.18	$7.59	0.00
CAB versus none	$6.84	$8.97	0.00	$6.63	$7.47	0.00	$6.72	$7.48	0.00
Marginal willingness-to-pay									
GT versus generic	$3.52	$4.23	0.20	$2.73	$2.87	0.35	$2.89	$3.10	0.38
Natural versus generic	$2.08	$3.75	0.06	$1.74	$2.34	0.13	$1.36	$2.12	0.12
Choice versus generic	$3.63	$4.97	0.04	$3.21	$3.22	0.51	$3.64	$3.58	0.50
CAB versus generic	$3.38	$5.20	0.02	$3.25	$3.38	0.25	$3.18	$3.47	0.37
Natural versus GT	−$1.43	−$0.47	0.16	−$0.99	−$0.54	0.15	−$1.53	−$0.98	0.11
Choice versus GT	$0.11	$0.74	0.21	$0.47	$0.35	0.77	$0.75	$0.47	0.69
CAB versus GT	−$0.14	$0.98	0.09	$0.52	$0.51	0.40	$0.28	$0.37	0.50
Choice versus natural	$1.55	$1.21	0.67	$1.46	$0.88	0.94	$2.28	$1.46	0.97
CAB versus natural	$1.29	$1.45	0.44	$1.51	$1.04	0.85	$1.82	$1.35	0.91
CAB versus choice	−$0.25	$0.24	0.19	$0.04	$0.16	0.16	−$0.46	−$0.11	0.28

[a]WTP values are derived from models in table 4. WTP values are for 12 oz steaks.
[b]Reported statistics are point estimate of WTP.
[c]p-Values report results of the one-sided test that the hypothetical WTP distribution > the nonhypothetical WTP distribution for each corresponding steak pair. p-Values were determined by applying the nonparametric combinatorial method (Poe, Giraud, and Loomis) to 1,000 bootstrapped WTP estimates calculated using the Krinsky–Robb bootstrapping method.

had a WARP violation. For those individuals that violated WARP, at most two violations were found (out of a possible 136 violations). Overall, across all individuals and possible comparisons, only 0.05% of nonhypothetical and 0.06% of hypothetical responses violated WARP. These results indicate that individuals in both treatments behaved consist with the weak axiom of revealed preferences across the seventeen CE scenarios. Even though participants in the hypothetical treatment were not held accountable for their choices, they behaved in a manner consistent with utility maximization that would imply stable preferences across choice sets.

Conclusions

This study compared hypothetical and actual choices in a study of consumer demand for quality-differentiated beef. The predicted probability of purchasing beef steaks was generally higher in the hypothetical versus non-hypothetical setting (as much as 30% higher for some steaks) and average willingness-to-pay to obtain steaks in the hypothetical setting was about 1.2 times that in the nonhypothetical setting. However, marginal willingness-to-pay to exchange steaks with differing quality attributes was, in general, equivalent in hypothetical and nonhypothetical settings. In fact, in two cases, marginal willingness-to-pay was significantly greater in the hypothetical than in the nonhypothetical setting. We also found that consumer preferences across the seventeen CE questions were as stable in the hypothetical context as when payment was actually required.

That we found evidence of hypothetical bias in a CE where Carlsson and Martinsson did not is likely attributable to a couple of factors. Carlsson and Martinsson's CE did not include a "none of these" option, which forced subjects to make a choice that included some level of payment. In contrast, we allowed subjects to opt out of making a purchase by choosing "none of these." In addition, Carlsson and Martinsson only analyzed marginal willingness-to-pay, whereas we analyzed both marginal and total willingness-to-pay. In fact, our results are consistent with those of Carlsson and Martinsson with regard to marginal willingness-to-pay.

For agribusiness applications, it should be straightforward to conduct nonhypothetical valuations following protocol outlined in this article, which, as we show, will likely yield different results than a hypothetical study. Understanding which value elicitation technique to employ and whether valuations should be conducted in a nonhypothetical context would certainly be enhanced by future research where comparable retail market-level data exist. The transportation literature comparing stated and revealed preferences has found that stated preferences provide reasonable estimates of marginal changes in quality attributes, but poorly predict actual market shares (Louviere, Hensher, and Swait). We observe a similar relationship between hypothetical and nonhypothetical responses to a CE, and in that regard, hypothetical CE responses may provide a good approximation to changes in market share and welfare associated with *marginal* changes in product quality.

References

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

Allenby, G.M., and J.L. Ginter. "The Effects of In-Store Displays and Feature Advertising on Consideration Sets." *International Journal of Research in Marketing* 12(1995):67–80.

Ben-Akiva, M., and S.R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press, 1985.

Bhat, C. "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research B* 29(1995):471–83.

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

Bradley, M., and A. Daly. "Use of the Logit Scaling Approach to Test for Rank-Order and Fatigue Effects in Stated Preference Data." *Transportation* 21(1994):167–84.

Carlsson, F., and P. Martinsson. "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?" *Journal of Environmental Economics and Management* 41(2001):179–92.

Carson, R.T., R. Groves, and M. Machina. "Incentive and Informational Properties of Preferences Questions." Plenary Address, European Association of Environmental and Resource Economists, Oslo, Norway, 1999.

Chalfant, J.A., and J.M. Alston. "Accounting for Changes in Tastes." *Journal of Political Economy* 96(1988):391–410

Cummings, R.G., G.W. Harrison, and E.E. Rutström. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review* 85(1995):260–66.

DeShazo, J.R., and G. Fermo. "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency." *Journal of Environmental Economics and Management* 44(2002):123–43.

Diamond, P.A., and J. Hausman. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives* 8(1994):45–64.

Erdem, T., and J. Swait "Brand Equity as a Signaling Phenomenon." *Journal of Consumer Psychology* 7(1998):131–57.

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

Greene, W.H. NLOGIT Version 3.0 Reference Guide. Econometric Software, Inc., 2002.

Hausman, J., and D.A. Wise. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46(1978):402–26.

Hensher, D.A., and M. Bradley. "Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models." *Marketing Letters* 4(1993):139–51.

Hensher, D., J. Louviere, and J. Swait. "Combining Sources of Preference Data." *Journal of Econometrics* 89(1999):197–221.

Jayne, T.S., L. Rubey, F. Lupi, D. Tschirley, and M.T. Weber. "Estimating Consumer Response to Food Market Reform Using Stated Preference Data: Evidence from Eastern and Southern Africa." *American Journal of Agricultural Economics* 78(1996):820–24.

Johnson, F.R., and W.H. Desvousges. "Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs." *Journal of Environmental Economics and Management* 34(1997):79–99.

Johnson, F.R., and K.E. Mathews. "Sources and Effects of Utility-Theoretic Inconsistency in Stated-Preference Surveys." *American Journal of Agricultural Economics* 83(2001):1328–33.

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

Krishnamurthi, L., S.P. Raj, and K. Sivakumar. "Unique InterBrand Effects of Price on Brand Choice." *Journal of Business Research* 34(1995):47–56.

Lancaster, K. "A New Approach to Consumer Theory." *Journal of Political Economy* 74(April 1966):132–57.

Layton, D.F., and G. Brown. "Heterogeneous Preferences Regarding Global Climate Change." *The Review of Economics and Statistics* 82(2000):616–24.

List, J.A. "Using Choice Experiments to Value Non-Market Goods and Services." Working paper, Department of Agricultural and Resource Economics, University of Maryland, 2001.

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

List, J.A., and J.F. Shogren. "Calibration of the Differences Between Actual and Hypothetical Valuations in a Field Experiment." *Journal of Economic Behaviour and Organization* 37(1998):193–205.

Louviere, J.J. "Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, and Recent Developments and External Validity." *Journal of Transport, Economics and Policy* 10(1988):93–119.

——. "Experimental Choice Analysis: Introduction and Overview." *Journal of Business Research* 23(1991): 291–97.

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

Louviere, J.J., and G. Woodworth. "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data." *Journal of Marketing Research* 20(1983):350–67.

Lusk, J.L. "Branded Beef: Is It What's For Dinner?" Choices 2nd Quarter 16(2001):27–30.

Lusk, J.L., and J.A. Fox. "Value Elicitation in Laboratory and Retail Environments." *Economics Letters* 79(2003):27–34.

Lusk, J.L., J. Roosen, and J.A. Fox. "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(2003):16–29.

Mackenzie, J. "A Comparison of Contingent Preference Models." *American Journal of Agricultural Economics* 75(1993):593–603.

McFadden, D. "Econometric Modes of Probabilistic Choice." In C. Manski and D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, MA: MIT Press, 1981, pp. 198–272.

McFadden, D., K. Train, and W.B. Tye. "An Application of Diagnostic Tests for Independence from Irrelevant Alternatives Property of the Multinomial Logit." *Transportation Research Record* 637(1977):39–46.

Poe, G.L., K.L. Giraud, and J.B. Loomis. "Simple Computational Methods for Measuring the Difference of Empirical Distributions: Application to Internal and External Scope Tests in Contingent Valuation." Staff Paper No. 2001-05. Department of Agricultural, Resource, and Managerial Economics, Cornell, University, September, 2001.

Poe, G.L., E. Severance-Lossin, and M. Welsh. "Measuring the Difference (X − Y) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76(1994):904–15.

Revelt, D., and K. Train. "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *The Review of Economics and Statistics* 80(1998):647–57.

Shogren, J.F., J.A. Fox, D.J. Hayes, and J. Roosen. "Observed Choices for Food Safety in Retail, Survey, and Auction Markets." *American Journal of Agricultural Economics* 81(1999):1192–99.

Swait, J., and W. Adamowicz. "The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching." *Journal of Consumer Research* 28(2001):135–48.

Swait, J., and J. Louviere. "The Role of the Scale Parameter in the Estimation and Use of Multinomial Logit Models." *Journal of Marketing Research* 30(1993):305–14.

Train, K.E. "Recreation Demand Models with Taste Differences Over People." *Land Economics* 74(1998):230–39.

Unterschultz, J., K.K. Quarainie, M. Veeman, and R.B. Kim. "South Korean Hotel Meat Buyers' Perceptions of Australian, Canadian, and US Beef. *Canadian Journal of Agricultural Economics* 46(1998):53–68.

Varian, H.R. "Non-Parametric Tests of Consumer Behaviour." *Review of Economic Studies* 50(1983):99–110.

——. "The Nonparametric Approach to Demand Analysis." *Econometrica* 50(1982):945–73.

Appendix

Choice Experiment Scenarios for 12-oz. Ribeye Steaks

Scenario	Generic Steak	Guaranteed Tender Steak	Natural Steak	USDA Choice Steak	Certified Angus Beef Steak
1	$3.38	$7.88	$6.75	$6.75	$6.75
2	$4.50	$9.00	$6.75	$5.63	$9.00
3	$6.75	$9.00	$5.63	$9.00	$6.75
4	$5.63	$9.00	$9.00	$6.75	$5.63
5	$6.75	$6.75	$6.75	$7.88	$5.63
6	$4.50	$7.88	$7.88	$9.00	$5.63
7	$4.50	$6.75	$5.63	$6.75	$7.88
8	$5.63	$5.63	$6.75	$9.00	$7.88
9	$5.63	$7.88	$5.63	$7.88	$9.00
10	$5.63	$6.75	$7.88	$5.63	$6.75
11	$6.75	$7.88	$9.00	$5.63	$7.88
12	$4.50	$5.63	$9.00	$7.88	$6.75
13	$3.38	$9.00	$7.88	$7.88	$7.88
14	$3.38	$5.63	$5.63	$5.63	$5.63
15	$3.38	$6.75	$9.00	$9.00	$9.00
16	$6.75	$5.63	$7.88	$6.75	$9.00
17	$5.63	$5.63	$5.63	$5.63	$5.63