Preferences With and Without Prices
- does the price attribute affect behavior in stated preference surveys?

Fredrik Carlsson[A]
Peter Frykblom[B]
Carl-Johan Lagerkvist[C]

Abstract

The experimental as well as the nonmarket valuation literature include several examples of how an introduced price can affect behavior in otherwise not expected ways. It has become standard to include a price vector as an attribute in choice experiments, something that enables us to estimate a marginal willingness to pay for other attributes. We test the impact on preferences by an inclusion of a price in a choice experiment. Preferences are affected, as might be expected. However, also the relative ranking of individual attributes is affected. We end on a positive note, observing that a price seems to drive out zero price opinions, e.g. warm glow values.

JEL classification: choice experiment, hypothetical WTP, preference reversal.

Key words: D12, D01, Q13

1. Introduction

Environmental and agricultural economics are two fields where hypothetical survey experiments are used extensively. In environmental economics, the focus is often on preferences for public good attributes, while in agricultural economics the focus is on preferences for private goods attributes that do not exist in the market. While contingent valuation (CV) experiments have been, and still are, the main applied methodology, there is increasing interest in choice experiments (CEs).[1] The fundamental difference between the two methods is, with CE, focus is on the valuation of separate attributes, while, with CV, the focus is on the valuation of certain combinations of attribute levels.

The issue that has attracted most attention, but also criticism, regarding CV is the validity of hypothetical responses; see for example Cummings *et al*. (1995), and List (2001). A few similar tests of CE have been undertaken, with mixed results; see for example Carlsson and Martinsson (2001) and Lusk and Schroeder (2004). In this paper, we raise a perhaps more fundamental question: are preferences over attributes affected by the presence of a price?

In a typical CE, the respondent is repeatedly asked to choose one of two or more alternatives. Each alternative is described by a number of attributes, whose levels vary across the choice sets. This feature allows us to estimate the marginal rate of substitution between attributes, and, if a price attribute is included, a marginal willingness to pay can also be estimated. Including the price attribute is thus critical from a welfare evaluation point of view. However, the role of the price attribute is potentially more than that. From the literature of preference reversals, we know that measuring preferences by choices or by prices can lead to different preference rankings (see for example Tversky and Thaler, 1990; Seidl, 2002; Slovic and Lichtenstein, 1983). While most preference reversals found in the literature entail different types of constructed or real-world lotteries, Haab and Roe (2003) find a reversal also for relatively simple work tasks with certain outcomes, and List (2002) finds it for different bundles of sports cards. Although a CE always requires a choice, it is still possible to test whether an included price attribute can affect the elicited preferences in a way resembling that in the preference reversal literature. In the CV literature, it is also well known the price vector can have anchoring effects that potentially affect the welfare estimates; see Green *et al*. (1998) and Frykblom and Shogren (2000) for discussions and

---

[1] For two reviews of choice experiments, see Alpizar, Carlsson, and Martinsson (2003) and Louviere, Hensher, and Swait (2000).

tests. Furthermore, there is evidence that even completely arbitrary number anchors can influence respondents (Ariely, Loewenstein and Prelec, 2003). CE, and other stated preference methods, often involve goods and attributes that can have non-use values; these choices can involve important ethical considerations, or what has been called "warm glow" (Andreoni, 1989). Stated preferences under such circumstances are perhaps more likely to be affected by the application of a price attribute.[2]

Intrigued by these different and partly confounded reasons why a price attribute can affect preferences over other attributes, we have undertaken two experiments that concern a choice between alternative food products, where the price is omitted in one of the experiments. Among the attributes included in the experiment, some can be expected to involve other than direct use values.

The rest of the paper is organized as follows. The choice experiment is described in section 2, followed by a presentation of the econometric method in section 3. In section 4, the results of the experiment and our tests are presented; section 5 concludes the paper.

2. The Choice Experiment

Altogether 1,600 surveys were sent out by mail. 800 received a traditional CE questionnaire where price was included as an attribute, while the remaining 800 individuals received a questionnaire not including any prices. All other aspects were identical in the two questionnaire versions. Each questionnaire included eight choices, four choices for the purchase of chicken fillets, and four for ground beef. A number of actions were taken in an attempt to design a survey that was policy relevant, plausible, and meaningful to the respondent. First, industry representatives and academic researchers who are specialized in farm animal production were consulted and involved in the process of devising the questionnaire. This was followed up by focus groups, where the participants were asked to complete a questionnaire and write down any questions and comments. The focus group participants later took part in a round-table discussion of the questionnaire. The results from the following pilot survey were returned to the individuals and organizations that participated. This was repeated three times until we had a satisfactory questionnaire. Each pilot study was distributed to a random sample of 200 individuals, and was conducted during May-September 2003.

---

[2] This can be compared with the research by Frey (1997), who shows how monetary incentives can crowd out intrinsic motivations and social norms.

The resulting questionnaire consists of three sections. The first includes questions about the respondent's and the household's habits regarding food consumption. The choice experiment constitutes the second section, and the third contains questions regarding the respondent's socio-economic and demographic status. In the introduction to the choice experiment, the purpose of the survey was briefly explained. This was followed by a description of the different attributes. The respondent was also provided with a separate fact sheet giving a description of each attribute. Two farm animal products were valued: chicken and ground beef (see the Appendix for an example of a choice set). Both questionnaires contained the following information (in this case applied to chicken)[3,4]:

> We now want you to choose between different types of chicken fillets. Every type of fillet is described with help of a number of attributes. We have attached a fact sheet where the attributes are explained in greater detail.
> **Please read through the fact sheet before you continue.**
> The price of chicken fillet varies depending on where you live and shop. A fillet costs on average SEK 80 per kilo. Your local price is most likely a few SEK more or less.

The following text was appended to the questionnaire that included a price attribute:

> We have changed the attributes of a chicken fillet in the alternatives that are presented on the next page, and often **raised the price**. If you now stand in the grocery store and choose between the different types of chicken fillets that are presented below, which one do you choose? In each of the four choice situations, we want you to choose one type of fillet, given the different prices. The given increases in price are not always realistic, but they help us understand what consumers value and want.

In the questionnaire that did not include a price, the following text was included in order to avoid respondents using their own guesses at reasonable prices for the alternatives:

> We have changed the attributes of a chicken fillet in the alternatives that are presented on the next page. If you now stand in the grocery store and choose between the different types of chicken fillet that are presented below, which one do you choose? In each of the four choice situations, we want you to choose one type of fillet. We ask you not to consider that

---

[3] All the forms, including the questionnaires, are available on request. Excerpts presented in the text are translations from the original questionnaires, which were printed in Swedish.
[4] At the time the surveys were carried out, SEK 7.5 was approximately equal to US $1.

production costs probably will be higher, so please choose one alternative and **assume that the price is the same** as you already pay today.

The choice sets were created using a cyclical design principle (Bunch, Louviere, and Andersson, 1996). A cyclical design is a straightforward extension of the orthogonal approach. First, each of the alternatives from a fractional factorial design is allocated to different choice sets. Attributes of the additional alternatives are then constructed by cyclically adding alternatives to the choice set, based on the attribute levels. The attribute level in the new alternative is the next higher attribute level to the one applied in the previous alternative. If the highest level is attained, the attribute level is set to its lowest level. Strictly dominating choice sets were deleted from the possible collection of choice sets. In addition, we wanted to avoid "too" dominant choice sets. This was done by calculating a so-called code sum for each alternative (Wiley, 1978). In order to calculate the code sum, we rank the levels of the attributes from worst to best, and the lowest attribute level is assigned the value 0, the next is 1, the next following 2 and so on. So, for a level four attribute, the highest value is 3. The code sum is the sum of all theses values for each alternative. By comparing the code sums, one can get a simple indication of which alternatives are particularly dominant. This is obviously a crude approach, and in order for it to work reasonably well, the utility difference between two levels should not be too different across attributes. In our case, we deleted all design alternatives with a code sum difference exceeding four; there were altogether 13 such design alternatives.

The attributes used in the choice experiments vary, as the relevant policy questions for different foods are not alike. Information about the attributes and the experiments in which they were included is presented in Table 1. Many of these attributes can be expected to involve both use and non-use values. For example, consumers may care about whether mobile slaughterhouses are used, if they believe the resulting meat quality to be better, but also because they care intrinsically about how the animals are treated.

Table 1. Attributes and levels

Attribute	Levels
Label (Beef)	1. Minimum required by law 2. Farm of origin and choice of animal husbandry
Fodder (Chicken and Beef)	1. No information whether or not GM fodder has been used 2. Label whether or not GM fodder has been used 3. Use of GM fodder is banned
Outdoor production (Chicken and Beef)	1. Herd kept outdoors summertime/Herd always kept indoors 2. Herd kept outdoors all year/Herd kept outdoors summertime
Transport (Chicken and Beef)	1. Transport of live animals to slaughterhouse 2. Mobile slaughterhouse
Growth (Chicken)	1. Fast-growth chicken (35-39 days) 2. Slower growth chicken (at least 81 days)
Cost (Chicken)	0 (80); +4 (84); +8 (88); +12 (92); +24 (104)
Cost (Beef)	0 (40); +4 (44); +8 (48); +12 (52); +24 (64)

3. Econometric specification

In the analysis of the responses, we apply a random parameter logit model (Train, 2003). Define a latent utility function of alternative i for individual q, at choice situation t, consisting of a systematic and a stochastic part:

$$V_{itq} = \beta' a_{it} + \varepsilon_{itq} \tag{1}$$

where a_i is the attribute vector, β is the corresponding parameter vector, and ε_{itq} is an error term. We assume that the parameter of all the attributes expect the price attribute, are independently normally distributed. This means for each attribute that we estimate a common parameter and an individual deviation in terms of a standard deviation parameter. We will assume the utility parameters vary among individuals, but are constant among the choice situations for each individual. This reflects an underlying assumption of stable preference structures for all individuals. If the $\varepsilon's$ are IID type I extreme value, we have a random parameter logit, or a mixed logit model. The probability - conditional upon a certain set of preference parameters - of alternative i for individual q in choice situation t is then the standard logit probability:

$$L_q(it \mid \beta) = \frac{\exp(\beta a_{it}/\tau)}{\sum_{j \in \mathbf{A}_t} \exp(\beta a_{jt}/\tau)} \tag{2}$$

where $\mathbf{A}_t = \{A_1,...,A_N\}$ is the choice set and τ is the so-called scale parameter. Note the estimated parameters are confounded with the scale parameter, and the scale parameter is related to the variance of the error term; see e.g. Louviere *et al*. (2000).[5]

4. Results

The sample was drawn from a population defined as persons between 18 and 75 years of age and having a permanent address in Sweden. In total, 697 (44%) individuals returned the questionnaire, of whom 41 were not available for analysis due to non-response to various questions. Table 2 presents the results for the random parameter logit models for the two choice experiments; with and without a price vector. The models were estimated with simulated maximum likelihood using Halton draws with 150 replications.[6]

---

[5] The variance of a logistic distribution is $\tau^2 \pi^2 / 3$. Hence, a large scale parameter is equivalent to a large variance and vice versa. For convenience the scale parameter is often set to one.
[6] See Train (2003) for details on simulated maximum likelihood and Halton draws.

Table 2. Coefficient estimates random parameter logit models (p-values in parentheses)

		Price attribute		No price attribute	
		Chicken	Beef	Chicken	Beef
	Parameters				
Label	Farm origin and choice of husbandry		0.478 (0.000)		0.665 (0.000)
Growth	Slower growth chicken	0.601 (0.000)		1.058 (0.000)	
Fodder	Use of GM fodder is banned	0.839 (0.000)	1.094 (0.000)	2.427 (0.000)	1.496 (0.000)
	Use of GM fodder has been labelled	0.433 (0.000)	0.336 (0.001)	2.137 (0.000)	1.125 (0.000)
Outdoor prod.	Outdoors all year / summertime	0.363 (0.000)	0.090 (0.248)	0.722 (0.002)	-0.023 (0.849)
Transport	Mobile slaughterhouse	-0.178 (0.013)	0.271 (0.000)	0.111 (0.564)	0.742 (0.000)
Price		-0.566 (0.000)	-0.566 (0.000)		
	Parameter standard deviations				
Label	Farm origin and choice of husbandry		0.728 (0.000)		1.376 (0.000)
Growth	Slower growth chicken	0.362 (0.118)		1.155 (0.006)	
Fodder	Use of GM fodder is banned	0.105 (0.804)	0.032 (0.971)	1.600 (0.002)	0.790 (0.044)
	Use of GM fodder has been labelled	0.362 (0.291)	0.101 (0.685)	1.735 (0.001)	0.377 (0.501)
Outdoor prod.	Outdoors all year / summertime	0.515 (0.002)	0.627 (0.000)	2.105 (0.000)	0.116 (0.782)
Transport	Mobile slaughterhouse	0.496 (0.004)	0.576 (0.000)	1.659 (0.000)	1.573 (0.000)

Most of the attribute parameters are significant, and the signs of the significant parameters are identical in the two experiments. Many of the parameter standard deviations are also significant. Using a likelihood ratio test, we reject the hypothesis of equal parameters in the two experiments, with and without a price vector, at the 1% significance level.[7] However, the pooling of two different datasets is problematical, as the estimated parameters are confounded with the corresponding scale parameters, the

---

[7] This is done by comparing each model with a simple naïve pooling of the two datasets.

τ in equation (2). One way of dealing with this problem is to adjust for a possible difference in scale between the datasets. This is done using the grid search procedure proposed by Swait and Louviere (1993) to estimate the *relative* scale parameter.[8] Let the relative scale parameter be $\mu = \dfrac{\tau^{No\,Price}}{\tau^{Price}}$. This means, if the relative scale parameter is less than one, that the variance is greater in the experiment with price attributes and vice versa. Using the grid search procedure, the relative scale parameter is estimated to be 0.77. When imposing this parameter on the data, we can still reject the hypothesis of equal preferences between the two treatments at the 1% level by applying a likelihood ratio test. Hence, we find a significant difference in preferences between the survey version, with and without the price vector, even when adjusting for possible differences in the variance of the responses which was also found to be greater in the experiment where price was included.[9]

So far we have only conducted a joint test of all the preferences for all attributes. For some attributes, it may be the case that the preferences do not vary between the two versions of the survey, while for others they do. As the coefficients of the two estimated models cannot be compared directly, due to possible differences in the scale parameter, we estimate the marginal rate of substitution (MRS) for each of the attributes. As the MRS is the ratio of two attribute parameters, the scale parameter will cancel out. The growth attribute is used as the numeraire for the chicken experiment, and the labelling attribute is the numerarie for beef. The standard errors of the MRSs are estimated with the Delta method. The results are presented in Table 3.

---

[8] When estimating the random parameter models with the grid search procedure, 30 replications were used instead of 150, as each random parameter model takes quite a long time to estimate.
[9] With the grid search we do not estimate any standard deviation of the relative scale parameter and we are thus unable to draw any firm conclusions about differences in variance, as the relative scale parameter may not be significantly different from one.

Table 3. Marginal rates of substitutions (standard errors in parentheses)

Chicken	Price vector		No price vector		
	MRS	Rank	MRS	Rank	P-value t-test
Growth	1	2	1	3	
GM: Ban	1.39	1	2.29	1	0.135
	(0.29)		(0.53)		
GM: Label	0.72	3	2.0	2	0.023
	(0.24)		(0.52)		
Out summer	0.60	4	0.68	4	0.794
	(0.17)		(0.26)		
Mobile	-0.30	5	0.10	5	0.066
slaughter	(0.12)		(0.10)		
Beef	MRS	Rank	MRS	Rank	P-value t-test
Labelling	1	2	1	3	
GM: Ban	2.29	1	2.25	1	0.956
	(0.43)		(0.52)		
GM: Label	0.70	3	1.69	2	0.071
	(0.26)		(0.48)		
Out all year	0.19	5	-0.03	5	0.367
	(0.l7)		(0.18)		
Mobile	0.57	4	1.12	4	0.088
slaughter	(0.16)		(0.28)		

Using a two-sided t-test to compare the MRS across treatments (last column, Table 3), the null hypothesis (the MRSs are equal) is rejected for the attribute GM: Label at the 5% significance level. At the 10% level, three additional attributes have different MRS in the two survey versions. Consequently, whether preferences – measured as the MRS - are affected by the price attribute varies among the attributes.

Even if preferences are affected by the inclusion of a price attribute, a less strict requirement is unaffected rankings of the attributes (in terms of preferences). The relative importance of the attributes, within each treatment, is revealed by the relative magnitude of the coefficients. Table 3 ranks the attributes for each product and each treatment. As we can see, there is one ranking reversal for each food type. However, only for chicken is the reversal significant, in the sense that the MRSs for the two attributes that change ranks differ significantly from each other in the treatment without a price vector (p-value = 0.06). Finally, a less strict test of the ranking is a comparison of ordinal rankings of attributes, with and without a price. We calculate the Spearman rank-order correlation coefficient, which indicates the degree of association between the rankings (Siegel and Castellan, 1988). An estimated correlation coefficient equal to 0.9 for both products means the null hypothesis of no association between the rankings is rejected for both beef and chicken at the 5% level.

5. Conclusions

Choice experiments and other stated preference methods are important tools for welfare analysis and to elicit preferences for various characteristics of market and nonmarket goods. To be able to conduct welfare analysis, a price attribute is necessary. An assumption made in mainstream economics is that the elicited preferences for individual attributes should be independent of whether the price attribute is included or not, or in other words that individuals have well-defined preferences.

Despite using a non-parametric test and few observations, we do find a strong correlation between the rankings of attributes. However, when applying stricter tests, we find one order reversal in each experiment, reject a null hypothesis of equal intensity for several of the preferences (measured in terms of marginal rates of substitution) and also reject a joint test of the same null hypothesis. Consequently, our results overall show quite distinct differences in preferences between the two experiments.

Nevertheless, our results beg the question: what preferences, if any, are valid? We do not have a simple answer to this, but the answer is important. Welfare analysis often involves large monetary trade-offs. Our results highlight a potential problem when preferences are elicited: the way in which preferences are decided can affect the results. For a similar conclusion applied to welfare analysis, see List *et al*. (2004). As discussed in the introduction, there may be several reasons for our results. In general, one could say that decision-making is a constructive process. In other words, preferences are, at least partly, constructed in the elicitation process; see Tversky (1996) for a penetrating discussion of this. In our case, we also believe that the types of values or attributes that are valued do matter. Also discussed in the introduction, "warm glow" values can arise when attributes are associated with potentially emotive ethical concerns. In our survey, such values would be related to the two GM attributes and Mobile slaughter. Our findings strongly support this hypothesis, as the inclusion of a price attribute often lowers the estimated MRS for these attributes. Despite using stated choices, instead of real economic commitments, prices do appear to drive out, at least to a degree, these "zero price" opinions. This last observation calls for further investigation.

References

Alpizar, F., Carlsson F. and Martinsson P. (2003). Using choice experiments for non-market valuation. *Economic Issues* 8: 83-110.

Andreoni, J. (1989) Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy* 97: 1447-1458.

Ariely, D., Loewenstein, G. and Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics* 118: 73-105.

Bunch, D., Louviere, J. and Andersson D. (1996). A comparison of experimental design strategies for choice-based conjoint analysis with generic-attribute multinomial logit models, Working Paper, Graduate School of Management, University of California, Davis.

Carlsson, F. and Martinsson P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? *Journal of Environmental Economics and Management* 41: 179-192.

Cummings, R.G., Harrison, G.W. and Rutström, E.E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice question incentive compatible? *American Economic Review* 85: 260-66.

Frey, B.S. (1997b) *Not Just for The Money. An Economic Theory of Personal Motivation*. Edward Elgar, Cheltenham.

Frykblom, P. and Shogren, J.F. (2000). An experimental testing of anchoring effects in discrete choice questions. *Environmental and Resource Economics* 16: 329-341.

Green, D., Jacowitz, K.E., Kahnemann, D. and McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20: 85-116.

Haab, T. and Roe, B. (2003). Are preferences ranks stable? The persistence of preference reversal asymmetries in a static choice setting. The Ohio State University, working paper: AEDE-WP-0032-03.

List, J.A. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *American Economic Review* 91: 1498-1507.

List, J.A. (2002). Preference reversals of a different kind: "The more is less" phenomenon. *American Economic Review* 92: 1636-43.

List, J.A., Berrens, R., Bohara, A. and Kerkvliet, J. (2004), Examining the role of social isolation on stated preferences. *American Economic Review* 94:741-752.

Louviere, J., Hensher D. and Swait J. (2000). *Stated Choice Methods*. Cambridge University Press, Cambridge.

Lusk, J.L. and Schroeder T.C. (2004). Are choice experiments incentives compatible? A test with quality differentiated beefsteaks. *American Journal of Agricultural Economics* 86: 467-82.

Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys* 16: 621-55.

Siegel, S. and Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill International Editions.

Slovic, P. and Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review* 83:156-161.

Swait, J. and Louviere, J. (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* 30: 305-314.

Train, K. (1998). Recreation demand models with taste differences over people. *Land Economics* 74: 230-39.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Tversky, A. and Thaler, R.H. (1990). Preference reversals. *Journal of Economic Perspectives* 5: 201-211.

Tversky, A. (1996). Rational theory and constructive choice. In: Arrow, K.J., et al., eds. *The Rational Foundations of Economic Behavior*: Proceedings of the IEA Conference held in Turin, Italy. *IEA Conference Volume*, no. 114. New York: St. Martin's Press; London: Macmillan Press in association with the International Economic Association: 185-97.

Wiley, J.B. (1978). Selecting Pareto optimal subsets from multiattribute alternatives. *Advances in Consumer Research* 5: 171-174.

Appendix

Choice 1, ground beef

Attributes (ground beef)	Ground beef 1	Ground beef 2
Label	Minimum required by law	Farm of origin and choice of animal husbandry
Fodder	Genetically modified products in fodder are forbidden	No information if genetically modified fodder has been used
Outdoor production	Outdoor summertime	Outdoor all-year around
Transport to slaughter	Mobile slaughter house	Transport of live animals
Price increase SEK/kg	+ SEK 4	+ SEK 8
(total cost)	(SEK 44)	(SEK 48)
Your choice (mark one alternative)		