How much is too much?

- An investigation of the effect of the number of choice sets, starting point and the choice of bid vectors in choice experiments

Fredrik Carlsson[A]
Peter Martinsson[B]

Working Papers in Economics no. 191
January 2006
Department of Economics
School of Business, Economics and Law
Göteborg University

Abstract
In a split sample design, we examine how the number of choice sets, design of the first choice set (starting point), and the choice of attribute levels in the cost attribute affect the precision in the elicited preferences in otherwise completely identical choice experiment surveys. These issues are investigated for Swedish households' marginal willingness to pay to reduce power outages. Our results indicate that neither the number of choice sets nor the starting point choice set has a significant impact on estimated marginal willingness to pay, while the effect was significant for the additive scaling of the cost vector. At the end of the paper we discuss the implications of our results on future developments and applications of choice experiments.

Key-words: Attribute levels; choice experiment; complexity; length; power outages; starting point.

JEL Classification: C25, C93, D12, Q41.

Financial support from Elforsk is gratefully acknowledged. The paper has benefited from comments by the members of the Market Design reference group.
[A] Department of Economics, School of Business, Economics and Law, Göteborg University, Box 640, 405 30 Göteborg, Sweden; Ph +46 31 7734174; Fax +46 31 773 10 43; E-mail fredrik.carlsson@economics.gu.se.
[B] (corresponding author) Department of Economics, School of Business, Economics and Law, Göteborg University, Box 640, 405 30 Göteborg, Sweden; Ph +46 31 7735255; Fax +46 31 773 10 43; E-mail peter.martinsson@economics.gu.se.

1. Introduction

Choice experiments have over the last 10 years or so become a more and more applied stated preference method, especially in transport, environmental and health economics. Compared with the contingent valuation method, however, the number of studies conducted is still relatively small. Thus, the number of studies that has focused on issues beyond obtaining the marginal rates of substitution between attributes, most commonly between a non-monetary and a monetary attribute expressed as marginal willingness to pay, is even smaller. One important issue when applying stated preference methods is if the answers by the respondents are sensitive to the framing of the actual choice situations, since there is considerable evidence that the framing of questions and the information provided in a survey may affect the answers (e.g. Ajzen et al., 1996; Kahneman et al., 1999). The other issue of concern is the elicitation of individuals' preferences per se. In a contingent valuation survey, individuals are asked either to state their maximum willingness to pay contingent on the scenario presented, i.e. an open-ended survey, or whether or not they are willing to pay a certain bid, i.e. a closed-ended format. In a choice experiment on the other hand, the individuals pick the preferred option in each of the choice sets presented to them. Each choice set consists of several alternatives and each alternative is described by several attributes. There are several steps in the development of a choice experiment study in addition to creating the scenario, which can broadly be divided into three groups: (i) determining which attributes to include, (ii) deciding the levels of the attributes and the number of choice sets and (iii) creating the choice sets by deciding which levels to include in each choice set, i.e. experimental design. The development of a closed-ended contingent valuation survey is less difficult compared to a choice experiment since only the cost attribute is included in the former. In this paper we concentrate on choice experiments and specifically on the second issue, deciding the levels of the attributes and the number of choice sets. We focus on three methodological issues: the number of choice sets, starting point and levels in the cost vector. These issues are investigated, based on a national sample in Sweden, for Swedish households' marginal willingness to pay for avoiding power outages of different durations starting either on weekdays or weekends separated by winter and non-winter seasons.

The length of the survey may, among other things, (i) affect the response rate, (ii) result in respondents using simpler decision rules such as ignoring certain parts of the information provided, (iii) facilitate learning and (iv) result in respondents responding more or less randomly especially in the end due to fatigue. From a theoretical point of view, the more choice sets that are answered by each respondent, the more information we obtain about their preferences ceteris paribus. Moreover, as argued by Louviere (2004), it is important not to have too few choice sets in the final design, since the objective of a survey is often to generalize the results beyond the choice sets presented in the survey. However, due to budget and logistical constraints, there is often an upper limit on the number of choice sets in the final design, which is affected by the sample size we can afford to survey and the number of choice sets that a respondent would like to and is able to answer. Thus, the natural question then is: How many choice sets are too many? In many applied studies, the number of choice sets in the final design as well as the number of choice sets presented to a respondent are both quite small,[1] and the stylized fact seems to be that respondents cannot handle too many choice sets. On the other hand, Louviere (2004, page 18) argues that: ”It is widely believed that ‘modeling’ individuals requires ‘smallish designs,’ but in contrast to the equivalent of widely held ‘academic urban myths’ in marketing and transport research, there is considerable evidence that humans will ‘do’ dozens (even hundreds) of T’s” (*T refers to choice sets; our remark*). Hensher et al. (2001) tested different numbers of choice sets (4, 8, 16, 24 and 32) for choosing between different flights, and they found small differences in the estimated elasticities among the versions with different numbers of choice sets. On the other hand, Hensher (2003) found that the number of choice sets had a statistically significant impact on the valuation of travel time savings.

In the contingent valuation literature, starting point bias has been frequently discussed. It means that in a closed-ended contingent valuation survey, there is positive correlation between the bid in a first question and the answer to a second question, which either could be another bid (closed-ended format) or maximum willingness to pay (open-ended format) (Cameron and Quiggin, 1994; and Herriges and Shogren, 1996). Thus, a test of starting point bias in a contingent valuation study is straightforward. However,

[1] From the final design, a number of blocks containing a number of choice sets from the final design are created, and each respondent answers one block only.

this situation is more complicated in a choice experiment with several attributes in an alternative. In order to test for a starting point bias, we create a choice set such that a large improvement can be obtained at a relatively low cost. If this affects behavior, i.e. results in starting point bias, we expect individuals to be less likely to choose expensive alternatives in the following choice sets compared to not having this choice set first ceteris paribus. In the closed-ended contingent valuation literature, a number of papers have explicitly discussed the issue of optimal bid vector design and the effect of the bid levels chosen on the precision in the estimated willingness to pay (e.g. Alberini, 1995). A closed-ended contingent valuation survey can in principle be seen as a choice experiment with one choice set. The choice of the optimal bids to include in a contingent valuation survey is less difficult since one of the two cost attributes is always zero, i.e. if individual says no to the bid then nothing is paid. A similar discussion on which attribute levels to attach to the cost attribute is relatively absent in the choice experiment literature, where the focus instead has been on creating the choice sets given that some levels have been selected. It should be noted that in many cases, the cost attribute is the only attribute to which levels can be chosen freely, since most of the other attributes will be assigned levels that are policy relevant and plausible, for example the number of hydroplants to build in a river or the health effects from reducing car traffic in the city center. We are not interested in the optimal design per se.[2] Instead, we investigate how the framing of the cost attribute affects behavior by applying an additive scaling, where a constant amount is added to the cost vector and this vector is then applied in another sub-sample. If the utility function is linear in cost, the scaling should not matter since it is only the difference in cost that matters. So a first question is, what happens if we assume a linear utility function, and would the two survey versions produce the same results? A second question is, how do they compare if we assume another functional form? The rest of the paper is organized as follows. In Section 2 we describe the survey on power outages among a random sample of Swedes and the survey versions designed in order to test our research questions. In Section 3 the results are presented and, finally, Section 4 concludes the paper.

[2] For discussions on optimal design see e.g. Carlsson and Martinsson (2003), Kuhfeld (2001), and Louviere et al. (2000).

2. The choice experiment survey

The choice experiment survey concerned Swedish households' marginal willingness to pay for reducing unplanned power outages. Each respondent made repeated choices between two different alternatives presented to them. The alternatives differed in terms of the number of outages separated on the time of the week (weekdays and weekend) and duration (4, 8 and 24 hours), as well as in terms of the cost for obtaining the stated level of outages. This means that together with the cost attribute, there are in total seven attributes. The numbers of power outages were presented as the amount of outages during the next five years separated on weekdays and weekends, and on the durations described above. Moreover, each respondent answered two different parts: the first considered outages during the winter (November-March), while the second outages during the rest of the year (April-October). Each part contained either six or twelve choice sets depending on the survey version.[3] Before they answered the choice experiment, the respondents were instructed to read a short scenario describing the attributes and some facts regarding power outages. The scenario presented to them is found in the Appendix. After the scenario, the respondents were asked to make repeated pair-wise choices. An example of a choice set is presented below in Figure 1.

>>> FIGURE 1

In order to investigate our research questions, we created four different survey versions. First we created a Base version. The survey was developed through a series of focus groups and a larger pilot study. In this Base version, we presented 12 choice sets to be answered by each respondent (6 choice sets per season), and the levels of the cost attribute were 125, 200, 225, 275 and 375 SEK.[4] Based on this Base version, we created three additional versions for methodological tests as presented in Table 1 below, where the number of choice sets or the levels of the cost attributes were changed.

>>> TABLE 1

[3] In the Starting point version, there are seven choice sets for the winter.
[4] At the time of the survey, 1 USD was approximately equal to 7.50 SEK.

In the Length version, respondent were asked to answer 24 choice sets instead of 12. In the Starting point version, an additional choice set was included and located first among the choice sets presented. The levels were set so that the respondents could obtain a large reduction in the number of outages at a low cost.[5] If this choice set is used as a starting point, then when answering the other choice sets, we expect that the respondents will be more likely to choose the alternative with the lower cost in a choice set ceteris paribus. Finally, in the Scope version, the levels of the cost attribute are shifted upwards by adding 200 SEK compared with the Base version. By assuming a linear utility function, the difference in costs is still the same, and thus the marginal willingness to pay should be unaffected. In Table 2 we summarize the other attributes and attribute levels used. The levels refer to the number of outages over the next five years. The reason why we used over the next five years was because we wanted to avoid describing the outages in fractions rather than using integer numbers.[6]

>>> TABLE 2

Since we do not focus on the experimental design of the choice sets per se, we base all the versions of the survey on the same design, but they are blocked differently depending on the number of choice sets asked to the respondents. The choice sets were created using a cyclical design principle (Bunch et al., 1996).[7] In addition, we wanted to

[5] The cost difference was only 75 SEK, and it resulted in a reducion of two 24 hour outages.
[6] In Sweden, households located in built-up areas experienced on average 0.08 planned and 0.39 unplanned power outages per year during the period 1998-2001 with an average duration of 12 and 23 minutes respectively (Svenska Kraftnät, 2002). The corresponding figures for power outages in sparsely populated areas of Sweden are 0.60 planned with average durations of 83 minutes and 1.54 unplanned power outages per year with an average of 203 minutes.
[7] 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 into the choice set based on the attribute levels of the first alternative. 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

avoid “too” dominating choice sets. This was done by calculating code sums for each alternative (Wiley, 1978).[8] From the remaining 584 choice sets, 36 choice sets were drawn with a linear D-optimal design. These were then blocked into 6 different blocks of choice sets, which were randomly allocated to the respondents. For the Length version, the sets were randomly allocated into 4 different blocks, each containing 9 choice sets. To the 9 choice sets we added 3 additional choice sets for both winter and the rest of the year in order to test for transitivity.[9]

The households’ choices can be described by using a random utility model framework. We assume a linear utility function, where the error term enters the function in an additive way. So the utility of alternative *i* is then

$$U_i = v_i + \varepsilon_i = \beta a_i - \gamma c_i + \varepsilon_{ij}\,, \tag{1}$$

where a_i is a vector of the attributes in alternative *i*, β is the corresponding parameter, c_i is the cost associated with alternative *i*, γ is the marginal utility of income and ε_i is an error term. If the error terms are extreme value distributed, then the probability that a particular alternative is chosen can be formulated as the standard logit probability. Thus, the probability that alternative *A* is chosen when there is a choice between A and B can be expressed as

$$P[A] = \frac{1}{1+\exp(-\beta(a_A - a_B) + \gamma(c_A - c_B))}\,. \tag{2}$$

In the literature, there is an increasing concern about the role of the scale parameter in discrete choice models and there are also increasing empirical evidences of the importance of modeling the scale parameter in an appropriate way. In particular, attributes may have effects both on the behavior in terms of affecting the level of the utility as well as in terms of affecting the variance of the utility. In order to assess the

---

level is set to its lowest level. Strictly dominating choice sets were deleted from the candidate set of choice sets.

[8] In order to calculate the code sum, we order the levels of the attributes from worst to best, with the lowest attribute level assigned the value 0, the next 1, the following 2 etc. So for a four level attribute the highest value is 3. The code sum is the sum of all these values for each alternative. By comparing the code sums one can get a rough indication of which alternatives are highly dominating. This is obviously a crude approach, and in order for the approach to work fairly 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 larger than eight; in total there were 78 such design alternatives in the candidate set.

[9] These three choice sets were constructed so that all three pair-wise combinations of three alternatives were compared. We did not include test of transitivity in the other versions, since only 3 choice sets would be from the original design in each part of the choice experiment. Since test of transitivity is not a focus of this paper, we do not discuss this test further, but a large majority passed the transitivity test.

potential effect on the level of utility, for example for a welfare analysis, it is important to make sure that one does not instead capture the effects on the variance; see for example Louviere et al. (2000), Islam and Louviere (2004) and Swait and Adamowicz (2001a,b). We therefore use a binary heteroskedastic logit model, where the error term has a logistic distribution with mean zero and variance $\exp(\delta z_i)^2$, where z_i is a vector of the choice set specific characteristics and δ is the corresponding parameter vector. The question remains what to include in the variance function. One obvious candidate is of course the attribute levels. However, since the variance function is exponential, it is not advisable to directly include the attribute levels directly in the variance function, because the discrete choice model depends on difference in attribute levels.[10] Therefore we include two other attribute characteristics of our choice experiment: the first one is the absolute value of the difference in total number of outages between the two alternatives, and the other is the absolute value of the difference in cost. Finally we also include a dummy variable for the first half of the total number of choice sets to test for differences in variance between the first and the second half of the experiment. Note that differences may occur either because respondents get tired or bored, or because of the fact that the first half of the experiment concerns outages during winter and the second half outages during the rest of the year.

Under the assumption of a linear utility function, the marginal willingness to pay for an attribute is the ratio between the parameter of the attribute and the cost parameter such that

$$MWTP = \frac{\beta}{\gamma}. \tag{3}$$

[10] For example, we would believe that the effect on the variance is the same for a choice set, where the cost of alternative A is 200 SEK and the cost of B is 250 SEK as for a choice set where the cost for A is 250 SEK and the cost for B is 200 SEK. However, since only the difference in attribute levels matters, the exponential variance function will not treat them as the same unless we use the absolute difference between them.

3. Results

The population that the sample was drawn from was defined as individuals aged between 18 and 74 years old with a permanent address in Sweden. A random sample of 3,600 individuals was then selected from the Swedish census registry. The mail survey was conducted in 2004 with one reminder sent out 10 days after the first questionnaire. The response rates on the different versions of our questionnaires are summarised below in Table 3.[11] It should be noted that the response rates only differed such that length resulted in a lower response rate.

>>> TABLE 3

To begin with we estimate separate binary heteroskedastic logit models for each of the three survey version. The results from these estimations are shown in Table 4.

>>> TABLE 4

As shown in the table above, all the utility parameters have the expected negative signs, i.e. the more outages or the higher the cost, the higher the disutility ceteris paribus. Thus, an increase in any of these attributes in an alternative in a choice set would decrease the probability of that alternative being chosen ceteris paribus. Moreover, most of the parameters are significant at the 1% significance level. The signs of the parameters in the variance function are the same across survey versions, while the level of significance differs to some extent among them. The coefficient of the variance function of the absolute cost difference is positive and significant at the 1% level in all versions. A positive sign on the coefficient implies that the variance increases as the difference in cost increases in a choice set. The interpretation is the same for the absolute difference in the hours of outages, but this effect is only significant in two out of the four versions. The coefficient of the dummy variable for the first half of the choice sets is negative, which indicates that the variance is higher for the second half of the experiment. This

[11] The response rates at the 1-digit postal code level corresponds to the population statistics at the same level, so the sample is representative with respect to geographic location. Comparing the sample statistics with population statistics shows no statistical difference at the 5% level related to gender composition, but a slight overrepresentation of older people (SCB, 2003, and SCB, 2004).

may indicate that (i) respondents get tired and/or bored while answering the survey, and/or (ii) they are more certain about their responses for the first half. Note that the first half of the experiment concerned outages during winter time, so this would mean that they are more certain about their preferences regarding outages during the winter.

We begin with testing the null hypothesis of equal parameters across survey versions. In order to test this we pool the Base version with each of the other versions one at a time, and estimate a pooled model. Using a likelihood ratio test we can then test the null hypothesis of equal parameters by comparing the separately estimated models with the pooled model. However, when pooling the data we need to control for a potential difference in the scale parameter, or the variance of the responses, among survey versions (e.g. Swait and Louviere, 1993; Louviere et al., 2000). This is done by estimating an additional variance term that considers the relative scale between the two survey versions.[12] We cannot reject the null hypothesis of equal parameters between the Base version and the Length version, or between the Base version and the Starting point version at the 5% significance level. However, we can reject the null hypothesis between the Base version and the Scope version at the 5% significance level. Hence, we concentrate on the marginal willingness to pay for the Base and the Scope versions.[13] Using the expression in equation (3) we see that the scale parameter cancels out. Thus, we can directly compare the calculated marginal willingness to pay between the two survey versions. The standard errors for these mean marginal willingnesses to pay are obtained by using the Delta method (Greene, 2000). The results are shown in Table 5 together with the results of a t-test of equality of marginal willingness to pay between the two survey versions.

For many of the outages there is a significant difference at the 5% significance level in marginal willingness to pay as shown in Table 5 column 4. Interestingly, the marginal willingness to pay is consistently higher for the survey version where the cost vector has been doubled. If there is an income effect, it would go in the opposite direction, since the income for the respondent is lower in the second version. In any case, the income

[12] All results are of course available upon request.
[13] We have costructed a similar comparison for the other versions as well, and for most of the marginal willingnesses to pay we cannt reject the hypotheses of equality.

effect should be small. Thus, the size of the cost vector does matter here, although it is only the difference in cost that should matter for their choice and the corresponding marginal willingness to pay. A possible explanation for this result is that respondents are affected by the size of the cost attribute apart from an income effect. Thus, the level of the cost attributes may work much like a signaling effect, where high cost levels signal to the respondent that one should pay more money.

4. Discussion

This paper reports on the results from a series of choice experiments focusing on the impact of the number of choice sets, starting point and the attribute levels in the cost attribute on the estimated marginal willingness to pay. By using the same experimental design, but blocked differently depending on the length of the survey, we hold the task complexity and the effect of design technique applied constant.

Our results are in line with the discussion in Louviere (2004), i.e. that respondents can answer many choice sets, which was supported empirically by Hensher et al. (2001). Thus, our results are encouraging since they indicate that many choice sets can be asked to each respondent, which benefits the possibilities to generalize the results beyond the choice sets included in the conducted survey by letting the choice sets from a relatively large final design be applied. However, the proportion of non-responders increased by almost 16% in the Length version which is significantly different from the Base version at the 1% significance level. However, the fraction of non-responses related to the introduction of more choice sets is low compared to the total non-response rate, although the non-response rate might be increased even further by giving more choice sets to each respondent. Thus, the purpose of increasing the number of choice sets is to encourage people to start answering. Once this hurdle is overcome, they have no problem answering more choice sets than what is claimed by “academic myths” (Louviere, 2004, p.18). In general, we would suggest the use of in-person interviews to both increase the proportion of responders as well as to increase the quality of the responses by being able to better communicate the scenario. One way to reduce the logistical problems with conducting interviews is to apply a stratified sampling (based on geographical areas).

Inclusion of a first choice set, where the respondents could for a relatively low cost obtain a relatively large improvement in terms of a reduced number of outages, had no significant impact on the estimated marginal willingness to pay. This finding suggests that the ordering of the choice sets does not influence the responses in the same way as found in closed-ended CV studies, and that a respondent treats one choice set as independent from another choice set. However, the level of the cost attributes had a significant impact on the marginal willingness to pay, suggesting that not only the difference between the levels of attribute is important, but also the absolute levels. In particular we found that the marginal willingness to pay is higher in the survey version with higher costs, although the cost differences were held constant. Note that if an income effect were present, we would have the opposite result. It is clearly difficult to draw any clear-cut conclusions regarding the design of future choice experiments from our results, not the least because the same effect may also occur in real life, i.e. the results are not only an artifact of stated preference surveys. On the contrary, one may expect similar effects for non-market goods in general, where individuals would use costs or bid levels as implicit willingness to pay cues.

References

Alberini, A. (1995) Efficiency v. Bias of Willingness-to-Pay Estimates: Bivariate and Interval Data Models, *Journal of Environmental Economics and Management* 29, 169-180.

Ajzen, I., T.C. Brown, L.H., Rosenthal and H. Lori (1997) Information Bias in Contingent Valuation: Effects of Personal Relevance, Quality of Information and Motivational Orientation, *Journal of Environmental Economics and Management* 30, 43-57.

Bunch, D., J. Louviere and D. Andersson (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.

Cameron, T. A. and J. Quiggin (1994). Estimation Using Contingent Valuation Data from a "Dichotomous Choice with Follow-Up" Questionnaire, J*ournal of Environmental Economics and Management* 27, 218-234.

Carlsson, F. and P. Martinsson (2003) Design Techniques for Stated Preference Methods in Health Economics, Health Economics 12, 281-294.

Greene, W. (2000) Econometric Analysis. New Jersey: Prentice-Hall.

Hensher, D.A., P.R. Stopher and J.J Jordan (2001) An exploratory analysis of the effect of number of choice sets in designed choice experiments: an airline choice application, *Journal of Air Transport Management* 7, 373-379.

Hensher, D.A. (2003) Revealing differences in willingness to pay due to dimensionality of stated choice designs: An initial assessment, Working Paper at Institute of Transport Studies, ITS-WP-03-08.

Herrings, J.A. and J.F. Shogren (1996) Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning, *Journal of Environmental Economics and Management* 30, 112-131.

Islam, T. and J. Louviere 2004. Modelling the Effects of Including/Excluding Attributes in Choice Experiments on Systematic and Random Components. Unpublished Working Paper.

Kahneman, D., I. Ritov and D. Schkade (1999) Economic Preferences or Attitudes Expressions?: An Analysis of Dollar Responses to Public Issues, *Journal of Risk and Uncertainty* 19, 203-235.

Kuhfeld, W. (2001) Multinomial Logit, Discrete Choice Modeling. An Introduction to Designing Choice Experiments, and Collecting, Processing and Analyzing Choice Data with SAS. SAS Institute TS-643.

Louviere, J.L. (2004) Random Utility Theory-Based Stated Preference Elicitation Methods: Applications In Health Economics With Special Reference To Combining Sources of Preference Data, Working Paper at Center for the Study of Choice No. 04-001.

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

SCB (2003), *Statistisk Årsbok För Sverige 2002* (Statistical Yearbook of Sweden 2002), Stockholm, Elanders Novum.

SCB (2004), *Statistisk Årsbok För Sverige 2003* (Statistical Yearbook of Sweden 2003), Stockholm, Elanders Novum.

Svenska Kraftnät (2002) Ett Robust Elförsörjningssystem, Svenska Kraftnät, Stockholm (in Swedish).

Swait, J. and W. Adamowicz (2001a) The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decisions Strategy Switching, *Journal of Consumer Research* 28, 135-148.

Swait J. and W. Adamowicz (2001b) Incorporating the Effect of Choice Environment and Complexity into Random Utility Models, Organizational Behavior and Human Decision Processes 86, 141-167.

Swait, J. and J. Louviere. (1993) The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* 30, 305-314.

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

Table 1. The different versions applied in the study.

Version	Number of choice sets	Cost attribute levels
Base	12	125, 200, 225, 275, 375
Length	24	125, 200, 225, 275, 375
Scope	12	325, 400, 425, 475, 575
Starting point	13	125, 200, 225, 275, 375

Table 2. Attribute and attribute levels in the choice experiment.

Attribute	Attribute levels
Number of outages of 4 hour duration over 5 years	2,1,0
Number of outages of 8 hour duration over 5 years	2,1,0
Number of outages of 24 hour duration over 5 years	1,0

Table 3. The response rates from different survey versions.

Survey version	Total sent	Answers	Proportion	Test of no difference in response rate compared to the Base version (p-value)[14]
Base	1200	470	39.2%	
Scope	800	321	40.1%	0.67
Starting point	800	316	38.3%	0.88
Length	800	265	33.1%	0.01

[14] We tested the null hypothesis of no difference in respose rate in the Base version and the other versions separately.

Table 4. Heteroskedastic logit models.

		Base version	Length version	Starting point version	Scope version
April-October	4 hour weekday	-0.167[a]	-0.199[a]	-0.151[a]	-0.084
	8 hour weekday	-0.299[a]	-0.211[a]	-0.251[a]	-0.450[a]
	24 hour weekday	-0.787[a]	-0.873[a]	-0.651[a]	-0.837[a]
	4 hour weekend	-0.200[a]	-0.303[a]	-0.236[a]	-0.352[a]
	8 hour weekend	-0.454[a]	-0.489[a]	-0.344[a]	-0.751[a]
	24 hour weekend	-1.015[a]	-1.178[a]	-0.973[a]	-1.071[a]
November - March	4 hour weekday	-0.122[b]	-0.198[a]	-0.145[b]	-0.197[a]
	8 hour weekday	-0.276[a]	-0.457[a]	-0.604[a]	-0.580[a]
	24 hour weekday	-1.031[a]	-1.208[a]	-0.966[a]	-1.385[a]
	4 hour weekend	-0.302[a]	-0.245[a]	-0.275[a]	-0.642[a]
	8 hour weekend	-0.534[a]	-0.675[a]	-0.654[a]	-0.692[a]
	24 hour weekend	-1.376[a]	-1.471[a]	-1.418[a]	-1.475[a]
	Cost	-1.667[a]	-1.881[a]	-1.506[a]	-1.175[a]
Variance function					
Abs(difference in outages hours)		0.167	0.744[a]	0.172	0.444[a]
First half of choice experiment		-0.273[a]	-0.255[b]	-0.317[a]	-0.393[a]
Abs(Cost difference)		0.458[a]	0.301[a]	0.417[a]	0.500[a]
Log likelihood		2499	1818	1761	2881

a Significant at the 1% level.
b Significant at the 5% level.

Table 5. Marginal willingness to pay with standard errors in parentheses.

	Base	**Scope**	
April-October	Marginal willingness to pay (standard error)	Marginal willingness to pay (standard error)	t-test (p-value)
4 hour weekday	10.0 (2.6)	7.1 (4.8)	0.5 (0.60)
8 hour weekday	17.9 (3.7)	38.3 (7.5)	2.4 (0.01)
24 hour weekday	47.2 (4.4)	71.2 (10.5)	2.1 (0.03)
4 hour weekend	12.0 (3.0)	29.9 (7.6)	2.2 (0.03)
8 hour weekend	27.2 (3.9)	63.9 (9.5)	3.6 (0.00)
24 hour weekend	60.9 (4.4)	91.1 (10.0)	2.8 (0.01)
November -March			
4 hour weekday	7.3 (3.4)	16.8 (6.6)	1.3 (0.20)
8 hour weekday	16.5 (4.7)	49.3 (10.4)	2.9 (0.00)
24 hour weekday	61.8 (6.3)	117.9 (17.3)	3.0 (0.00)
4 hour weekend	18.01 (4.2)	54.6 (12.0)	2.9 (0.00)
8 hour weekend	32.0 (5.2)	58.9 (13.1)	1.9 (0.06)
24 hour weekend	82.5 (6.8)	125.5 (16.7)	2.4 (0.02)

Figure 1. Example of a choice set.

Number of outages during 5 years	*Alternative A*	*Alternative B*
Working days: number of outages with 4 hour duration.	0 during 5 year	1 during 5 year
Working days: number of outages with 8 hour duration.	1 during 5 year	0 during 5 year
Working days: number of outages with 24 hour duration.	1 during 5 year	0 during 5 year
Weekends: number of outages with 4 hour duration.	1 during 5 year	2 during 5 year
Weekends: number of outages with 8 hour duration.	0 during 5 year	1 during 5 year
Weekends: number of outages with 24 hour duration.	0 during 5 year	1 during 5 year
Connection fee for your household.	200 SEK	225 SEK
Your choice		

Appendix. Scenario

We will now ask some questions regarding your household's willingness to avoid power outages. Imagine that there is the possibility of choosing among different contracts with your electricity supplier and that a backup electricity board exists that can be used in case of a power outage. By connecting to this backup electricity board you can reduce the number of power outages that your household will experience. For connection to this service you have to pay a connection fee to the owner of the network. Apart from the stated power outages there will always be a number of power outages that you know about in advance because of the need to conduct maintenance work.

Since power outages are not particularly common, we present the number of outages for a 5 year period. For each alternative we will state the number of power outages of varying durations on working days (Monday-Friday) and weekends (Saturday–Sunday). The time of the year when the power outage occurs may have an impact on your experience of the power outage. We will therefore ask questions both for power outages during winter and during the rest of the year.

An example of a choice set is shown below. For each set we want you to state which alternative you think is best for you and your household. Note that your choice will not affect anything other than the number of power outages and your fixed tariff - everything else remains as it is today.

[An example of a choice set was shown below].