Accepted Manuscript

Examining ordering effects in discrete choice experiments: A case study in Vietnam

Thanh Cong Nguyen, Jackie Robinson, Shinji Kaneko, Nguyen The Chinh

PII: S0313-5926(15)00005-3
DOI: http://dx.doi.org/10.1016/j.eap.2015.01.003
Reference: EAP 49

To appear in: *Economic Analysis and Policy*

Received date: 25 August 2014
Revised date: 27 January 2015
Accepted date: 27 January 2015

Please cite this article as: Nguyen, T.C., Robinson, J., Kaneko, S., Chinh, N.T., Examining ordering effects in discrete choice experiments: A case study in Vietnam. *Economic Analysis and Policy* (2015), http://dx.doi.org/10.1016/j.eap.2015.01.003

***Title Page [include author details]**

Examining ordering effects in discrete choice experiments: A case study in Vietnam

Thanh Cong Nguyen[a,]*, Jackie Robinson[b], Shinji Kaneko[c], Nguyen The Chinh[d]

[a] Faculty of Environment and Urban, National Economics University, Hanoi, Vietnam

[b] School of Economics, University of Queensland, Brisbane 4072, Australia

[c] Graduate School for International Development and Cooperation, Hiroshima University, Higashi-Hiroshima 739-8529, Japan

[d] Institute of Strategy and Policy on Natural Resources and Environment, Ministry of Natural Resources and Environment, Hanoi, Vietnam

* Corresponding author. Tel.: +84 24 36280280, ext: 5137 (office); +84 944008982 (Mobile)

Fax: +84 24 38698231

E-mail addresses: thanh.nguyen9@uqconnect.edu.au (Thanh C. Nguyen), robinson@uq.edu.au (J. Robinson), kshinji@hiroshima-u.ac.jp (S. Kaneko), ntchinh@isponre.gov.vn (Nguyen T. Chinh).

Examining ordering effects in discrete choice experiments: A case study in Vietnam

ABSTRACT

The order of a series of choice tasks presented to respondents in a discrete choice experiment (DCE) could affect the choice outcomes. This study explores the ordering effects in a DCE surveying preferences for improvements in cyclone warning services in Vietnam. Respondents' choices are analysed non-parametrically and parametrically to investigate the ordering effects in their preferences. Across the sequence of six choice questions, the stated demand of respondents is statistically significantly different at the first position from all other positions. Based on a parametric analysis using mixed logit models, we also find that the willingness-to-pay for a number of improvement programs estimated at the first position is relatively larger when compared with the other positions. The findings indicate that although DCEs can provide additional information on respondents' preferences when compared with survey methods using a single valuation question, the trade-off for more information is the ordering effects over a sequence of repeated questions.

KEYWORDS

Ordering effect; Discrete choice experiment; Stated preference; Vietnam; Developing country

JEL CLASSIFICATION CODES

C81, L97, Q51

1. Introduction

Discrete choice experiments (DCEs) have been increasingly used to value a range of multi-attribute public goods and services (Hoyos, 2010). In a standard DCE survey, each respondent is requested to make choices in repeated valuation tasks, such that more information on respondents' preferences can be collected from each DCE survey relative to a survey containing a single valuation question. However, the additional information elicited by the repeated-question format is challenged by the body of evidence for ordering effects (Day et al., 2012; Day and Pinto Prades, 2010; Holmes and Boyle, 2005; Ladenburg and Olsen, 2008; McNair et al., 2011; Scheufele and Bennett, 2012), which contends that stated preferences may change when the valuation questions are presented in a different order.[1] In relation to the ordering of choice tasks, Day et al. (2012) suggest that there are two main types of ordering effects: (1) position-dependent ordering effect and (2) precedent-dependent ordering effect. The first represents the changes in respondents' preferences relating to the position of a choice task in the series of choice tasks. The second type of ordering effect refers to the changes in respondents' stated preferences relating to features of the choice alternatives in previous choice tasks, which can be the first choice task or the best or worst option in the range of previous choice tasks.

The position-dependent ordering effect casts doubt on the standard assumption that preferences are stable across a sequence of discrete choice questions (Day et al., 2012; Day and Pinto Prades, 2010; McNair et al., 2011). This type of ordering effect presents stated preference (SP) practitioners with a serious issue that responses in a series of repeated choice questions might not tell us about 'true' preferences when respondents' choices appear to change when the choice questions are presented in a different position.

---

[1] Ordering effects may also be related to the ordering of attribute presented in choice tasks (Farrar and Ryan, 1999; Kjær et al., 2006). However, this attribute ordering effect is out of scope of the present paper.

The precedent-dependent ordering effect, however, "should not necessarily be taken as evidence of some inherent problem" with using the repeated choice question format in a DCE exercise (Day et al., 2012) p.89. The marketing literature has demonstrated that consumers' purchasing decisions are based on reference prices, which are shaped by consumers' prior experience and current purchase environment (Mazumdar et al., 2005). Putler (1992) claims that the reference price effects can be empirically tested, which was achieved in a study using weekly retail egg sales data from Southern California. Isoni (2011) also suggests that the effects of best or worst deal on purchasing decisions appear intuitively appealing because of their resemblance to the experience of everyday transactions. For example, the feeling of disappointment at knowing that a product we just bought can be found for a cheaper price is a common experience (Isoni, 2011). Nevertheless, Day et al. (2012) suggest that the presence of the precedent-dependent ordering effects should be addressed since it is closely related to the issue of *strategic misrepresentation* of preferences in which a respondent might provide untruthful answers so as to manipulate the survey outcome to his/her benefit. For instance, respondents might reject improvement options with cost levels higher than the lowest cost level previously observed in the sequence of choice tasks in order to encourage low cost of provision in the future. The precedent-dependent ordering effects, therefore, may affect willingness-to-pay (WTP) values estimated from a series of choice questions.

There is a number of previous studies which have addressed the ordering effects in DCE exercises undertaken in developed countries (Carlsson et al., 2012; Day et al., 2012; Day and Pinto Prades, 2010; McNair et al., 2011; Scheufele and Bennett, 2012, 2013). To examine changes in respondents' preferences in a sequence of choice tasks, previous studies have adopted a number of approaches: comparing single-question and repeated-question split samples (McNair et al., 2011; Scheufele and Bennett, 2013), presenting the same blocks of choice tasks in different orders (Carlsson et al., 2012; Day et al., 2012), and repeating the same choice tasks in different orders (Day and Pinto Prades, 2010; McNair et al., 2011; Scheufele and Bennett, 2012). While all the previous studies have examined issues related to the position-dependent ordering effects, some of the studies have addressed topics related to the precedent-dependent ordering effects (Day et al., 2012; Day and Pinto Prades, 2010; McNair et al., 2011; Scheufele and Bennett, 2012). When investigating the precedent-dependent ordering effects, Day et al. (2012) took into account both cost and non-monetary attributes by calculating a 'deal' value for every alternative; the other studies limited the scope of their examination to only the cost attribute (Day and Pinto Prades, 2010; McNair et al., 2011; Scheufele and Bennett, 2012).

Findings about changes in respondents' preferences across a sequence of choice tasks are mixed. In a sequence of 16 choice tasks, where the last block of 8 tasks are identical to the first block, Carlsson et al. (2012) found that respondents became less cost sensitive in the later block, resulting in higher WTP estimates for the second block. Comparing three blocks of 5-6 choice tasks, parametric models in the Day et al. (2012) study, suggest that respondents seemed to maintain a constant cost sensitivity in the sequence of choice tasks, and that there might be a position-dependent decrease in WTP driven by an increasing tendency to choose the status-quo option but not through a decreasing assessment of the value of non-monetary attributes. In the other studies, results show that respondents' cost sensitivity increased along the sequence of choice tasks, so that WTP estimates decreased in the later positions or in the repeated-question split samples when compared with the single-question format sample (McNair et al., 2011; Scheufele and Bennett, 2012, 2013). Previous studies have shown that the changes in respondents' preferences along a sequence of choice tasks have a number of possible explanations, such as learning effects, strategic misrepresentation of preferences, income uncertainty effects, fatigue effects and reference effects (Carlsson et al., 2012; Day et al., 2012; Day and Pinto Prades, 2010; McNair et al., 2011; Scheufele and Bennett, 2012, 2013).

Despite the differences in research methods, scope and results, previous studies provide robust evidence of ordering effects, which do not support a naive interpretation of choices in repeated-question format surveys as revealing information on the assumption of stable ‘true’ preferences. The DCE approach to valuation has been increasingly applied to address problems in developing countries (Bennett and Birol, 2010; Mangham et al., 2009). The experiences from the DCE literature (Bennett and Birol, 2010; Cook et al., 2007; Do and Bennett, 2008; Mangham et al., 2009; Othman et al., 2004; Tuan and Navrud, 2007) suggest that respondents in developing countries are capable of taking part in a SP survey and that their responses are generally reasoned and deliberate. However, applying the DCE method in the developing country context faces some particular challenges, such as respondents' lack of experience with SP surveys and/or a low level of literacy (Bennett and Birol, 2010; Mangham et al., 2009). Given the challenges, ordering effects would be a serious issue in the application of the DCE method in the developing country context. For example, respondents in developing countries, who are unfamiliar with SP surveys and/or have a low level of literacy, might be prone to progressive fatigue when responding to a series of choice questions; and the fatigue may increase the tendency to choose the status-quo option considered as a safe and easy choice in the later choice tasks.

To complement previous studies in the literature, all undertaken in the developed country context, we seek to provide empirical evidence for the ordering effects in the context of a developing country (i.e. Vietnam). Our case study provides insight into ordering effects in a developing country, which would be useful for DCE practitioners in designing their DCE applications in the developing country context. In this paper, a non-parametric analysis is undertaken based on investigating changes in respondents' stated demand along a sequence of choice tasks. A parametric analysis using mixed logit models is constructed to examine both position- and precedent-dependent ordering effects. In the next section, we introduce our DCE exercise conducted in Vietnam and present the research methods in more detail. Results of our non-parametric and parametric analyses are reported in Section 3. The final section presents concluding remarks.

2. Research design and Method

2.1. Overview of the DCE survey

A recent DCE study aimed at estimating the WTP for improvements in tropical cyclone warning services in Vietnam is used as the basis for this paper (Nguyen et al., 2013). With a geographical position close to the tropical cyclone centre of the western North Pacific, Vietnam is one of the most cyclone-prone countries in the Mekong region. In country rankings on cyclone fatal risk, Vietnam is among the top 20 countries in the world (Mosquera-Machado and Dilley, 2009; Peduzzi et al., 2012). In Vietnam, general public forecasts and warnings of severe weather and climate events (i.e. tropical cyclone warnings) are freely provided and usually disseminated to communities through mass media such as television and radio. This information enables communities to assess the risk of an approaching cyclone and to respond to threats from the tropical cyclone. In recent years, the number of powerful cyclones affecting Vietnam has increased (Nguyen et al., 2013). The warming climate could be responsible for the increased severity of tropical cyclone risk in Vietnam (MONRE, 2009). In 2010, the Government of Vietnam ratified the Strategy for Development of Hydro-meteorological Service until 2020, which is expected to improve the capacity of meteorological agencies. The improvements would have positive effects on the reduction of cyclone-related causalities and property damage in Vietnam.

The DCE exercise reported here included successive rounds of design and testing. The design of the DCE required identification of attributes or characteristics of meteorological services and the levels to be offered. Identification of appropriate attributes for inclusion in the DCE was based on information from previous studies (Gunasekera, 2004; Lazo and Chestnut, 2002; Lazo and Waldman, 2011; Lazo et al., 2010), verified for their suitability for Vietnam by meteorological experts in Vietnam and tested through an internet survey and several face-to-face interviews in a number of coastal areas of Vietnam. The three attributes adopted for this DCE exercise were accuracy of forecast information, frequency of update and mobile phone short message warning. In relation to the accuracy of cyclone forecast information, focus group discussions indicated that users are primarily concerned with the accuracy of projected location and timing of landfall (Nguyen et al., 2013). In this DCE exercise, the accuracy attribute was presented as an improvement between levels 1 to 3 relative to the current condition, which were explained to respondents using the descriptions presented in Appendix A. WTP was estimated by including a cost attribute in the choice tasks. In this study, the electricity bill was the vehicle for a one-off mandatory payment. The appropriateness of the selected attributes and the questionnaire were tested in two pilot surveys. Table 1 presents the levels of the attributes applied in this DCE exercise. Nguyen et al. (2013) provide a more detailed discussion about the design of the survey.

The next step was to construct a series of choice tasks to be presented to respondents using a choice task design. To minimize the correlation between the attribute levels in choice tasks, this study used an orthogonal design to generate twenty-four choice tasks. The twenty-four choice tasks were divided into four blocks of six choice tasks using the blocking procedure in *Ngene*[2], so that each respondent randomly answered a block of six choice tasks. This blocked design was devised to reduce the cognitive burden of respondents. In each choice task, respondents were requested to indicate their preference between two alternatives: one potential improvement program and the status quo (which kept all attributes at their current levels). The status quo option was identical across all choice tasks. An example of a choice task presented to respondents during this DCE is given in Figure 1. After the last choice task, respondents were requested to indicate whether they had ignored each of the four specified attributes when making their choices by answering Yes/No to four follow-up questions.

In 2011, face-to-face surveys were undertaken of 1133 household representatives at four sites representing both urban and rural coastal communities located in Northern and Central regions of Vietnam (Figure 2). The surveys were implemented in close collaboration with the National Centre

[2] *Ngene* is a software package developed for choice experiment design by ChoiceMetrics (http://choice-metrics.com).

for Hydro-meteorological Forecasting (NCHMF) and the Institute of Strategy and Policy on Natural Resources and Environment (ISPONRE), which are units within the Ministry of Natural Resources and Environment (MONRE). The surveys were implemented as a project of the ISPONRE. The credibility of the survey was assured by a ISPONRE recommendation letter and village leaders accompanying interviewers who were hydro-meteorological staff members. Carson and Groves (2007) suggest that respondents' belief in the consequentiality of the survey is one of the important conditions for eliciting truthful answers from respondents. For this survey, the collaboration with government agencies (ISPONRE and NCHMF) was expected to provide this positive influence.

As can been seen in Table 2, the total number of household representatives, who completed the questionnaire was 1014, providing a response rate of 89%. Table 3 presents a summary of the socio-economic characteristics of the sample in our DCE exercise in comparison with the characteristics of the Vietnamese population[3]. Statistical tests of differences in the characteristics between the sample and the Vietnamese population show that significant differences exist in most characteristics. The first characteristic that significantly differs from that of the Vietnamese population is the percentage of male respondents. A possible explanation is that men are usually responsible for tropical cyclone preparedness (i.e. strengthening house). Besides, the survey sites were in coastal areas where fishing is a common occupation, and fishing is also traditionally an occupation for men. It appears that men pay more attention to cyclone warning services, so they are more likely to represent their households to respond to the cyclone-related interview. An alternative possible reason is that in Vietnam, especially rural areas, men are usually head of household; therefore men often play the role as household representative to answer such questionnaires. A high percentage of male respondents might be a good indication that households in the sample were serious about their involvement in the survey, since the most relevant household member was chosen to participate.

[3] The socio-economic characteristics of the Vietnamese population were collected from the 2009 Vietnam Population and Housing Census and Vietnam Household Living Standards Survey (VHLSS) 2010.

Table 3 also shows that the average age of respondents in the sample is significantly higher than that of the Vietnamese population. A reason may be that older people who have more experience of tropical cyclones are more likely to be willing to participate in the interview. Monthly household income collected in our survey is lower than that presented in the Vietnam Household Living Standards Survey 2010. This is because the survey sites are in coastal areas and the poverty level in coastal areas of Vietnam is high (PEP, 2008). Poorer households in coastal areas of Vietnam are more likely to have a larger number of members, since they cannot afford a new house for younger couples in the family. The household size in the surveyed sample is larger than the equivalent number for the Vietnamese population. The sample for the present study, which targeted coastal areas, is likely to be representative of the population most affected by cyclone risk in Vietnam. Unfortunately, the authors could not find any studies providing information on the socio-characteristics of cyclone risk affected populations in Vietnam. Since the sample is not representative of all households in Vietnam, care should be taken when interpreting the results on a population level.

2.2. Examination of ordering effects

- *Non-parametric analysis*

In our DCE survey, each block of six choice tasks, from which respondents made their choice, were presented using a series of six showcards; the order of the showcards was varied randomly for each respondent. This means that each choice task was presented to respondents at different positions. This design allows us to examine how respondents' stated demand for an improvement option in each choice task may change along the six positions. Our non-parametric analysis of position-dependent ordering effects is undertaken based on the examination of acceptance rates (proportion choosing an improvement option over the status quo option) along the positions. Friedman's ANOVA, which is a non-parametric test, is applied to test for any significant difference in the acceptance rates calculated for the twenty-four choice tasks along the six positions. Friedman's ANOVA tells us only that a difference exists; it does not show specifically where the difference lies. To have a clearer picture of the differences in the acceptance rates along the six positions, several Wilcoxon signed-rank tests[4] are also conducted.

- *Parametric analysis*

The parametric analysis of ordering effects in our DCE exercise is based on the application of mixed logit (ML) models (Hensher and Greene, 2003; Train, 2009). In a ML model, the utility (U_{ikt}) associated with each alternative k, as evaluated by each individual i in choice task t, can be approximated by a linear function form as follows:

$$U_{ikt} = \alpha_{ik} + \beta_i{}'X_{ikt} + \varepsilon_{ikt} \quad (1)$$

where α_{ik} is the coefficient on an alternative specific constant (ASC) representing the utility associated with moving away from the status quo option, β_i is the vector of taste parameters, X_{ikt} is the vector of independent variables that are observed by the researcher, and ε_{ikt} is the stochastic unobserved component. In each choice task, respondents are assumed to choose an option that yields a better utility.

In the parametric analysis, the position-dependent ordering effects are investigated using interactions of five dummy variables, representing five positions from 2 to 6, with attribute variables (i.e. accuracy, updating frequency, mobile phone short message warning, and cost). Position 1 is used as the baseline and has a coefficient of 0, such that the other five position dummy variables are assessed relative to position 1. Interactions of the position dummy variables with the ASC are also included in the parametric econometric model to address possible changes in utility associated with the status quo option along the series of choice tasks.

[4] Wilcoxon signed-rank test, which is also a non-parametric test, is used to compare two related groups.

Following Day et al. (2012), the 'deal' value for each improvement option was calculated to explore the precedent-dependent ordering effects. The deal values represent respondents' assessments or impression about benefits from the improvement options. While the status quo option has a deal value of zero, the deal value for the improvement options is defined as the relative improvements in the quantity of cyclone warning attributes, i.e. accuracy, frequency of update and mobile phone short message warning, per a unit of cost (Day et al., 2012). A respondent i, who reported that he/she ignored or attended to attribute j, may think an improvement option k has a deal value of v_{ik} as follows:

$$v_{ik} = \frac{\sum_j \left(\beta_j / min\beta_j\right) z_{ij} X_{jk}}{X_{cost,k}} \qquad (2)$$

where β_j is the taste parameter of attribute j (i.e. accuracy, frequency of update and mobile phone short message warning) and is estimated from a ML model including only the attribute variables; z_{ij} = 0 if the respondent i reportedly ignored the attribute j and 1 if otherwise; X_{jk} is the change in quantity of attribute j in an improvement option k relative to the status quo option; $X_{cost,k}$ is the cost level in option k. As seen in the formula for deal value, the taste parameters, β_j, are used to form a preference-weighted sum of the improvement quantity of the warning attributes in order to reflect differences in respondents' preferences (Day et al., 2012). The extension of Day et al. (2012)'s formula in our study is to take into account respondents' statements on attribute ignoring, such that dummy variables for attribute ignoring, z_{ij}, are included to calculate the preference-weighted sum.

In this DCE exercise, a series of six choice tasks each of which consisted of the status quo and an improvement option were presented to respondents. In the sequence of choice tasks, respondents might assess an improvement option by comparing its deal value with the deal values of improvement alternatives presented in the previous choice tasks. The previously observed choice tasks may include the best/worst option in the range of previous choice tasks, the first choice task, and the immediately preceding choice task (Day et al., 2012). To examine these precedent-dependent ordering effects, four variables were constructed: *best deal*, *worst deal, first deal, and* *previous deal*. For each choice task, while the differences across choice tasks in the deal value of the status quo are zero, the variable values related to the improvement option are calculated accordingly as follows:

- Value of *best deal* variable is the difference between the deal value of improvement option in the current choice task and the maximum deal value in the series of previous choice tasks, excluding the first choice task.
- Value of *worst deal* variable is the difference between the deal value of improvement option in the current choice task and the minimum deal value in the series of previous choice tasks, excluding the first choice task.
- Value of *first deal* variable is the difference between the deal value of improvement option in the current choice task and the deal value of improvement option in the first choice task.
- Value of *previous deal* variable is the difference between the deal value of improvement option in the current choice task and the deal value of improvement option in the immediately preceding choice task.

- *Investigation of willingness-to-pay estimates*

One interesting research question is how the ordering effects lead to changes in WTP estimates along the sequence of choice tasks. Answers to this research question would provide a better understanding of ordering effects in our study. Changes in WTP estimates show the simultaneous changes in preferences for both the cyclone warning attribute and cost across the sequence of choice questions. The WTP for each warning attribute *j* is calculated as follows:

$$wtp_j = -\frac{(\beta_j + \beta_{position*warning\ attribute\ j})}{(\beta_{cost} + \beta_{position*cost})} \tag{3}$$

where β_j = parameter estimate of warning attribute j, β_{cost} = cost attribute parameter estimate, $\beta_{position*warning\ attribute\ j}$ = parameter estimate of interaction between position variables and warning attribute j, and $\beta_{position*cost}$ = parameter estimate of interaction between position variables and cost.

3. Results

3.1. Results of non-parametric identification

An analysis of the acceptance rates was conducted to examine changes in the stated demand of respondents across the six positions (P1, P2, P3, P4, P5, P6). The experimental design resulted in a total of twenty-four choice tasks, and the acceptance rates were calculated for the improvement option in each of the twenty-four choice tasks. The acceptance rates for the choice tasks across the six positions are reported in Appendix B. The mean acceptance rate is 50% for P1, 38% for P2, 34% for P3, the same rate of 32% for P4, P5 and P6.

There exist differences in the acceptance rates calculated for the twenty-four choice tasks across the six positions. The acceptance rates were tested for differences across the six positions using Friedman's ANOVA. The p-value estimated from Friedman's ANOVA is $p < 0.001$; therefore, we can reject the null hypothesis of no difference. Several Wilcoxon signed-rank tests were conducted to develop a clearer picture of the differences in the acceptance rates along the six positions. Table 4 presents p-values from the Wilcoxon signed-rank tests. For example, p-values in the second column of Table 4 show us whether there are significant differences in acceptance rates between position 1 and each of other positions (positions 2 to 6).

However, using several Wilcoxon signed-rank tests as a follow-up step for Friedman's ANOVA may inflate the Type I error rate[5] (Field, 2005). The important implication of the inflated Type I error is that researchers are more likely to falsely reject the null hypothesis of no differences (no ordering effects), leading to an overstatement of ordering effect issue in DCEs. The critical level of significance applied in Wilcoxon signed-rank tests following up Friedman's ANOVA should be adjusted to ensure that Type I errors for the group of tests do not increase to more than standard significance levels (e.g. 0.05). Field (2005) suggests the simplest method is to use a *Bonferroni correction*. This means that instead of using 0.05 as the critical level of significance, researchers should use a critical value of 0.05 divided by the number of tests conducted. The number of tests should be selective to avoid using a critical value that is too small, and therefore, too restrictive. The mean acceptance rates for P4, P5 and P6 are the same, hence, these three groups are considered as one group. The number of groups under consideration was reduced to four (P1, P2, P3 and P4-5-6), the number of Wilcoxon signed-rank tests, therefore, should be six. The critical value for significance that should be applied here is 0.01 or 1% (~ 0.05/6). Inspection of Table 4 reveals that at the 1% level of significance, the acceptance rate of P1 is statistically significantly different from all other positions, and the acceptance rate of P2 is only significantly different from P5 and P6. At 1% level, statistically significant differences were not found in the acceptance rates in the other positions.

[5] Imagine a situation in which there were three groups of respondents, and we were interested in differences between these three groups. If we were to carry out Wilcoxon signed-rank tests on every pair of groups, then we would have to conduct three separate tests. If each of these tests uses a 5% level of significance, then for each test the probability of falsely rejecting the null hypothesis (known as a Type I error) is only 5%. If we assume that three groups are independent, then the overall probability of no Type I errors for three Wilcoxon signed-rank tests is $(0.95)^3 = 0.95 \times 0.95 \times 0.95 = 0.857$. Now the probability of making Type I errors for three tests is 1 - 0.857 = 0.143 (14.3%). While the probability of making Type I error for one test is 5%, the equivalent probability for three independent tests is 14.3%.

3.2. Results of parametric models

The non-parametric analysis indicates that significant differences in the stated demand for improvements in cyclone warning services exist between the first position and other positions in the series of choice questions. However, it does not show us whether the ordering effects result in statistically significant changes in respondents' preferences for a specific attribute. The parametric analysis with the use of ML models seeks to clarify this question.

A ML model with the ordering-related variables was estimated using *NLOGIT 5.0*. In the ML model, the frequency of update and cost attributes are treated as continuous variables, while the attributes of accuracy improvement levels and mobile phone short message warning are modelled as dummy variables. Accuracy levels 2 and 3 are assessed relative to accuracy level 1; and after accounting for the ASC, the implied coefficient of accuracy level 1 is 0. The ML model has coefficients on each attribute, except for the cost attribute, specified as random parameters with normal distribution. The ML model also allows free correlation among the random parameters.

Table 5 presents the results of the ML model that has an acceptable pseudo-R^2 of 0.36. Mean coefficients of attribute variables, except for frequency of update, are statistically significant. These mean coefficients also have a positive sign as expected. The cost coefficient is statistically significant at the 1% level, and is negative which is consistent with economic theory. The estimated standard deviations, which are all statistically significant, represent unobserved heterogeneity in preferences for the related attributes.

With regard to the position-dependent ordering effects, the interactions between the position variables with the ASC are all statistically significant at the 1% level. In the ML model, the ASC parameter is confounded with the dummy accuracy parameters (please refer to Appendix C for more detailed discussion). Due to the confounding and the exclusion of the parameter of accuracy level 1, the ASC represents not only the utility of moving away from the current situation but also the preference for accuracy improvement level 1. The negative sign of the position interactions implies that when compared with position 1, respondents are less likely to choose improvement alternatives and options containing accuracy improvement level 1 in the later positions. As presented in Figure 3, the changes relative to position 1 in the acceptance rates of both improvement options and options containing accuracy level 1 show a decreasing trend. In positions 4 and 6, there are larger drops in the acceptance rate of options offering accuracy level 1, resulting in a larger interaction coefficient with the ASC for positions 4 and 6 relative to the other positions. This implies that the findings in the parametric model are consistent with the results of the non-parametric analysis. Moreover, the decreasing trend in choosing improvement alternatives, or in other words, the increasing tendency to choose the status quo along the sequence of choice tasks in the present study is similar to the finding of Day et al. (2012).

In the ML model, *accuracy level 2* and *accuracy level 3* are assessed relative to accuracy level 1. The variables of *accuracy level 2* and *accuracy level 3* represent the preferences for improvements from accuracy level 1 to accuracy levels 2 and 3. The position-dependent variables interacting with *accuracy level 2* and *accuracy level 3* show how respondents are likely to choose accuracy improvement level 2 and level 3, respectively, over accuracy level 1 in the later positions when compared with position 1. In position 2, the position-dependent variables interacting with both *accuracy level 2* and *accuracy level 3* are statistically significant and negative. This indicates that respondents are less likely to change from accuracy level 1 to opt for accuracy level 2 and accuracy level 3; in other words, respondents maintained their choice of accuracy level 1 in position 2, given that there were three accuracy improvement levels 1, 2 and 3 presented in the improvement alternatives. As can be seen in Figure 3, the change relative to position 1 in the acceptance rate of options containing accuracy level 1 is zero in position 2, indicating that the acceptance rate of options containing accuracy level 1 in position 2 is not different from the equivalent rate in position 1.

Across most of the other positions, respondent's preferences for improvements from accuracy level 1 to accuracy levels 2 and 3 are stable. Only in position 6 is the interaction with accuracy level 3 statistically significant. The negative sign of this interaction implies that respondents are less likely to switch their choice from accuracy level 1 to accuracy level 3. This finding is consistent with the changes in the acceptance rates presented in Figure 4. In position 6, there is a decrease in the acceptance rate of options containing accuracy level 1, but there is not an increase in the acceptance rate of options offering accuracy level 3. To some extent, the results of the parametric models are consistent with the results of the acceptance rates in non-parametric analysis. However, a limitation in the parametric model is that it is not able to identify some potentially significant changes in the acceptance rates presented in Figure 4 (e.g. the changes in position 4). A reason for this limitation could be the confounding between the ASC and the dummy accuracy parameters. Because of this confounding, the ML model is not able to separately measure the preference shifts from the status quo level of accuracy to accuracy improvement levels 1, 2 and 3, which are represented by the changes in acceptance rates in Figure 4.

In relation to position dependence of other attributes of cyclone warning services, inspection of the ML model presented in Table 5 reveals that all the interactions between the position-dependent variables and the attributes of frequency of update and mobile phone short message warnings are not statistically significant. The results show that respondents' preference for the frequency of update and the mobile phone message warning service is stable across the sequence of choice questions.

Concerning the cost attribute, the position-dependent variables interacting with this attribute are statistically significant in positions 3 to 6. The positive sign of these significant interactions indicates that respondents are more likely to choose an option offering higher levels of the cost attribute, or in other words, are less sensitive to the increased cost in the later positions relative to position 1. To gain a better understanding of respondents' choices over three cost levels (i.e. 50, 150 and 250 thousand VND) applied in this research, the acceptance rates of improvement options offering different levels of the cost attribute across the sequence of choice tasks are investigated. As shown in Figure 5, across the sequence of choice tasks, the acceptance rate of the cost level of 50,000VND is always highest; the cost level of 150,000VND has the second highest acceptance rate; and the lowest acceptance rate belongs to the highest cost level of 250,000VND. This confirms our expectation that the acceptance rate of an improvement option decreases as the increased cost in the household electricity bill for supporting that option rises. Investigation of Figure 6 reveals that from position 4, the acceptance rate of options containing the cost level of 150,000VND show an increasing trend. In positions 4 and 6, while there are sharp decreases in the acceptance rate of options offering the cost level of 50,000VND, the acceptance rate of alternatives offering cost level 250,000VND shows marked increases. The changes in the acceptance rates of higher cost levels are consistent with the parametric model results of position interaction terms with the cost attribute in positions 3-6.

The finding of a decrease in cost sensitivity in the later positions contradicts the increase in cost sensitivity found in McNair et al. (2011) and Scheufele and Bennett (2012). An explanation may be that McNair et al. (2011) and Scheufele and Bennett (2012) focused on examining respondents' strategies which tend to be driven mainly by the cost attribute. In their parametric models, only the cost attribute was specified to interact with position-dependent variables; and other factors, especially the constant that represents unobserved utility associated with the status quo, were assumed to be stable over a series of choice tasks. With this model specification, the position variables interacting with the cost attribute would attempt to capture all changes across the sequence of choice tasks, including a possible increasing proclivity to choose the status quo option. Therefore, it is likely that the coefficients of position interactions with the cost attribute would have a negative sign, indicating the increasing cost sensitivity. In our parametric model, not only the cost attribute but also all other attributes and the ASC were interacted with the position-dependent variables to examine changes in preferences associated with all the attributes and the ASC. With the differences in the model specifications, it is reasonable that our results of position interaction terms with the cost attribute contradict those found in McNair et al. (2011) and Scheufele and Bennett (2012). However, the finding of the decreasing cost sensitivity in this research is also different from the finding of Day et al. (2012), who found that respondents seem to exhibit a constant assessment of marginal disutility of cost across a sequence of choice questions.

There are a number of possible reasons for the decreasing cost sensitivity in this DCE data. A possible explanation is the issue of strategic misrepresentation of preferences. Some respondents who realise that they will not actually have to pay their stated WTP, may be more likely to choose higher cost levels in the later choice questions. An alternative reason for the decreasing cost sensitivity may be the interviewer biases. Given that the cost levels could be low,[6] some respondents may attempt to please the interviewers, who were hydro-meteorological staff members, by showing their WTP at the higher levels of cost in the later choice questions.

In relation to precedent-dependent effects, the *first deal* variable is statistically significant at the 1% level, confirming the importance of the deal value of an improvement option in the first choice task (Day et al., 2012). With the significance of w*orst deal* variable, the minimum deal value in the sequence of previous choice tasks also is important in shaping respondents' preferences. The coefficients of these significant precedent-dependent variables are positive, indicating that the deal value of an improvement option in the current choice task will be considered more favourably if it is greater than the deal value of the first choice and the minimum deal value in the previously observed deal values. The *best deal* and the *previous deal*, which are not statistically significant in the parametric model, do not have noticeable effects on respondents' preferences for cyclone warning services.

[6] The highest level of 250 thousand VND appears not to be sufficiently high to choke off demand for the good under concern. Whittington (1998) suggests that the highest level of cost "should be rejected by 90-95 percent of the respondents." (page 24). In our survey, about 18% of respondents still accepted the highest cost level.

3.3. Changes in willingness-to-pay estimates

To give a clearer pattern of how the WTP estimates change across the sequence of choice tasks, all the value estimates are reported in the graphical form in Figure 7. Our investigation of WTP estimates starts with total WTP estimated for two improvement programs:

+ Medium improvement program includes accuracy level 2, update frequency of 12 times and mobile phone short message warning.

+ Maximal improvement program includes accuracy level 3, update frequency of 16 times and mobile phone short message warning.

The total WTP values estimated for the medium and maximal improvement programs have a decreasing trend across the series of six choice tasks (Figure 7.a1). This is consistent with the decreasing pattern in the acceptance rate of improvement alternatives presented in Figure 7.a2. Tests using the resampling approach (Poe et al., 1997) confirm that the total WTP estimates in the first position are significantly higher than the equivalent amounts in the other positions at the 1% level of significance.[7] Using 1% as the critical level of significance, no significant differences were found in the total WTP estimates in positions 2 to 6. Our finding of higher total WTP estimates in the first position is consistent with recent findings in the literature (Day et al., 2012; McNair et al., 2011; Scheufele and Bennett, 2012).

While McNair et al. (2011); Scheufele and Bennett (2012) find that the reason for their higher WTP estimates for the first choice task is an increasing cost sensitivity, the finding of Day et al. (2012) shows an increasing tendency to choose the status quo over improvement alternatives. In our DCE data, the fall in the total WTP estimates could be driven by an increasing tendency to choose the status quo option (and not through an increasing cost sensitivity). A possible reason for the

[7] The p-value matrix of the resampling tests is reported in Appendix D.

increasing tendency to select the status quo option over improvement alternatives is respondent's uncertainty about the costs of improvement programs. Carson and Groves (2007) note that the presentation of different levels of the cost attribute may alert respondents to the prospect that the price they might have to pay is uncertain. Provided that this uncertainty is perceived as uncertain changes in their future income, risk averse respondents may exhibit a lower WTP by choosing the status quo option containing the level of zero cost.

An alternative reason for the increasing likelihood of choosing the status quo is that respondents become progressively fatigued when making choices in a series of repeated choice tasks. In this DCE exercise, respondents were requested to answer eleven questions, including questions on assessments of the current cyclone warning service and their exposure to cyclone risk, before reaching the choice questions. These eleven questions could have created a significant cognitive burden on respondents. If a respondent was becoming fatigued when making decisions on the sequence of six choice tasks, they could increasingly opt for the "safe choice" offered by the status quo option (Day et al., 2012).

In relation to marginal WTP estimates, the parametric analysis shows that respondents maintain a stable preference for both frequency of update and the mobile phone based warning service. Inspection of the resampling tests presented in Appendix D also reveals that there is no significant difference at the 1% level in the marginal WTP for these two attributes. The acceptance rates of options containing specific levels of updating frequency are relatively stable over the sequence of choice tasks (Figure 7.b2). In positions from 2 to 5, the updating frequency level 16 always has the highest rate of acceptance, and the acceptance rate of options containing the medium level of 12 times per day is higher than the equivalent rate of level 8. Concerning the mobile phone based warning service, the acceptance rate of options having this attribute is relatively low in the first position and highest in the last position (Figure 7.b2). This pattern of the acceptance rate of options offering the mobile phone based warnings is similar to their marginal WTP presented in graphical form in Figure 7.b1.

Regarding the accuracy attribute, the trend observed in the WTP estimates for the accuracy levels does not match the pattern of acceptance rates of options containing equivalent levels of accuracy improvement. The WTP estimates for the accuracy improvement levels presented in Figure 7.c1 exhibit a decreasing trend similar to the total WTP estimates for the medium and maximal improvement programs, while the acceptance rates of options containing the accuracy levels have a different pattern (Figure 7.c2). An explanation for this situation is the limitation of the confounding between the ASC parameter and the accuracy parameters (this issue has been discussed at length in Appendix C). Due to this confounding, the ASC parameter is used to estimate the WTP for the improvement levels of accuracy. A significant component of the preferences associated with the ASC is the utility associated with moving away from the status quo. By including the utility associated with moving away from the status quo, the WTP estimates for accuracy level 1, level 2 and level 3 should be interpreted as total WTP values for a program with only the accuracy attribute improved up to level 1, level 2 and level 3, respectively. Hence, the WTP estimates for the accuracy levels have a similar trend to the total WTP estimates for the medium and maximal improvement programs. The resampling tests, presented in Appendix D, also confirm that the WTP estimates for the accuracy levels are significantly higher in the first position at the 1% level of significance.

In the DCE data for this research, there is an increasing tendency to choose the status quo option over improvement alternatives across the sequence of choice tasks. Due to the use of the ASC in the estimation of the WTP for the accuracy levels, these WTP estimates decrease across the sequence of choice tasks to reflect the increasing propensity of choosing the status quo option associated with the ASC. The tendency to choose the status quo option appears to dominate changes in preferences for the accuracy levels, making it difficult to examine preference shift across the sequence of choice tasks for the accuracy levels based on their WTP estimates.

Given the consistency between the acceptance rates for the other attributes with their marginal WTP estimates, the acceptance rates of options containing the accuracy levels are reasonable to estimate a sense of change in preferences for the accuracy improvement levels across the sequence of choice tasks. Compared with the stable pattern of the acceptance rates of options containing the given levels of updating frequency, changes in the acceptance rates for the accuracy improvement levels presented in Figure 7.c2 appear to be relatively unstable. In the first three positions, the proportions choosing options containing accuracy level 1 are highest. In position 3, the acceptance rate of options offering accuracy level 3 is lowest. These findings are not consistent with what could be expected that the highest level of improvement should be the most accepted level. Only from position 4 onwards, the proportion of respondents choosing options containing accuracy level 3 is stable and highest. In positions 4 and 6, the acceptance rates of options offering accuracy level 1 are lowest as expected. This situation may be commensurate with the theory of *institutional learning*. Braga and Starmer (2005) suggest that respondents would learn more about the choice context, and the offered good as they progress through a series of choice questions. Given that the accuracy attribute was described to respondents as a complex composite of errors in the forecasts of cyclone position and landfall time,[8] this attribute would be difficult for respondents to understand and make an appropriate choice between the three levels. It appears that only from position 4 onwards have respondents been more familiar with the choice context and have a better understanding of the three levels of accuracy improvement.

4. Concluding remarks

The order of a series of choice tasks presented to respondents in DCEs could affect the choice outcomes. The data from the DCE survey in Vietnam are used to examine the ordering effects within the sequence of six repeated choice questions. In line with previous studies, this study finds the presence of both position- and precedent-dependent ordering effects in the DCE data. Across the sequence of six choice questions, the stated demand of respondents is statistically significantly

[8] Please refer to Appendix A for the description of the accuracy attribute and its levels.

different at the first position from all other positions. Results of the parametric analysis also show that the deal value of an improvement option in the first choice task is important in shaping respondents' preferences. The first choice effect appears to result in relatively larger estimates of total WTP for a number of improvement programs (i.e. the medium and maximal improvement programs, the programs with only the accuracy attribute improved up to level 1/level 2/level 3) estimated at the first position when compared with the other positions. While the finding of higher WTP estimates in the first position is contrary to the Carlsson et al. (2012) study, it is consistent with findings from other previous studies examining the stability of preferences (Day et al., 2012; McNair et al., 2011; Scheufele and Bennett, 2012, 2013). Despite the mixed results on the direction of changes in WTP estimates, the body of evidence accumulated thus far shows that WTP estimates from a sequence of choice tasks are significantly different from those estimated from the incentive compatible single binary choice task (McNair et al., 2011; Scheufele and Bennett, 2013). When compared with survey methods using a single valuation question, DCEs can provide additional information on respondents' choices, but a trade-off for more information is the ordering effects along a sequence of repeated questions. It is suggested that ordering effects should not be ignored when WTP results are used to inform decision-makers (Carlsson et al., 2012; Day et al., 2012).

Our findings also show some useful implications for the design of DCEs. Firstly, our examination of the acceptance rates suggests that respondents seemed to formulate a stable response strategy from position 4 onwards. The mean proportion choosing an improvement option over the status quo option was at the same rate (32%) for positions 4-6. In relation to the accuracy attribute with the three complex levels, it appears that from position 4 onwards respondents became familiar with the three levels of accuracy improvement, so that they made choices consistent with common sense, that the highest level of improvement should be the most accepted level. Hence, as suggested by Ladenburg and Olsen (2008), DCE practitioners should consider including examples of a few choice tasks, not generated by the experimental design and which are not intended for inclusion in the econometric analysis. It is likely that this approach would help to familiarise respondents with the choice context, and thus reduce potential ordering effects. Secondly, the low cost levels in our DCE survey could be a possible explanation for the decreasing cost sensitivity along the sequence of six choice tasks in our DCE data. Previous studies investigating ordering effects also note that cost levels play an important role in shaping respondents' preferences in a sequence of choice tasks. To mitigate ordering effects, it is highly recommended that more attention is paid to the selection of cost levels in the design of a DCE. Finally, we cannot rule out the possibility that fatigue effects could be an explanation for the tendency to opt for the status quo option in the later choice tasks. Given the fact that respondents in developing countries lack experiences with SP surveys and have a low level of literacy, fatigue effects should not be neglected in the choice task design. It suggests that DCE practitioners should carefully choose the number of choice tasks presented to each respondent in order to reduce the potential threat of ordering effects to DCE validity in the developing country context.

Acknowledgements

This study is a part of a research project “Estimating the benefits of an improved tropical cyclone warning service in Vietnam: An application of choice modelling” carried out with the aid of a grant from the Economy and Environment Program for Southeast Asia (EEPSEA). The authors are grateful to Vic Adamovicz, University of Alberta; Pham Khanh Nam, University of Economics Ho Chi Minh City for their valuable comments and suggestions on the research proposal and analysis. The authors would like to thank Le Thanh Hai, Truong Tuyen for their assistance in conducting the surveys.

References

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

Adamowicz, W., Dupont, D., Krupnick, A., Zhang, J., 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. Journal of Environmental Economics and Management 61, 213-226.

Alpízar, F., Carlsson, F., Martinsson, P., 2001. Using Choice Experiments for Non-Market Valuation. Economic Issues 8, 83-110.

Bennett, J., Adamowicz, V., 2001. Some fundamentals of environmental choice modelling, in: Bennett, J., Blamey, R.K. (Eds.), The choice modelling approach to environmental valuation. E. Elgar Pub., Cheltenham, UK ; Northampton, MA, USA, pp. 37-69.

Bennett, J., Birol, E., 2010. Choice experiments in developing countries: implementation, challenges and policy implications. Cheltenham, UK ; Northampton, MA : Edward Elgar.

Blamey, R.K., Bennett, J.W., Louviere, J.J., Morrison, M.D., Rolfe, J., 2000. A test of policy labels in environmental choice modelling studies. Ecological Economics 32, 269-286.

Boxall, P., Adamowicz, W.L., Olar, M., West, G.E., Cantin, G., 2012. Analysis of the economic benefits associated with the recovery of threatened marine mammal species in the Canadian St. Lawrence Estuary. Marine Policy 36, 189-197.

Boxall, P.C., Adamowicz, W.L., Swait, J., Williams, M., Louviere, J., 1996. A comparison of stated preference methods for environmental valuation. Ecological Economics 18, 243-253.

Braga, J., Starmer, C., 2005. Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis. Environmental and Resource Economics 32, 55-89.

Carlsson, F., Mørkbak, M.R., Olsen, S.B., 2012. The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling 5, 19-37.

Carson, R., Groves, T., 2007. Incentive and informational properties of preference questions. Environmental and Resource Economics 37, 181-210.

Cook, J., Whittington, D., Canh, D.G., Johnson, F.R., Nyamete, A., 2007. Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. Economic Inquiry 45, 100-114.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P., 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63, 73-91.

Day, B., Pinto Prades, J.-L., 2010. Ordering anomalies in choice experiments. Journal of Environmental Economics and Management 59, 271-285.

Do, N.T., Bennett, J., 2008. Estimating Wetland Biodiversity Values: A Choice Modelling Application in Vietnam's Mekong River Delta. Environment and Development Economics 14, 163-186.

Farrar, S., Ryan, M., 1999. Response-ordering effects: a methodological issue in conjoint analysis. Health Economics 8, 75-79.

Field, A.P., 2005. Discovering statistics using SPSS, 2nd ed. London ; Thousand Oaks, Calif. : Sage Publications.

Gunasekera, D., 2004. Economic Issues Relating to Meteorological Provision. Bureau of Meteorology research centre: Melbourne, Australia.

Hensher, D., Greene, W., 2003. The Mixed Logit model: The state of practice. Transportation 30, 133-176.

Holmes, T.P., Boyle, K.J., 2005. Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions. Land Economics 81, 114-126.

Hoyos, D., 2010. The state of the art of environmental valuation with discrete choice experiments. Ecological Economics 69, 1595-1603.

Isoni, A., 2011. The willingness-to-accept/willingness-to-pay disparity in repeated markets: loss aversion or ‘bad-deal’ aversion? Theory and Decision 71, 409-430.

Kjær, T., Bech, M., Gyrd-Hansen, D., Hart-Hansen, K., 2006. Ordering effect and price sensitivity in discrete choice experiments: need we worry? Health Economics 15, 1217-1228.

Ladenburg, J., Olsen, S.B., 2008. Gender-specific starting point bias in choice experiments: Evidence from an empirical study. Journal of Environmental Economics and Management 56, 275-285.

Lazo, J.K., Chestnut, L.G., 2002. Economic value of current and improved weather forecasts in the US household sector. Report prepared for the National Oceanic and Atmospheric Administration. Stratus Consulting Inc. Boulder, CO. November 2002.

Lazo, J.K., Waldman, D.M., 2011. Valuing improved hurricane forecasts. Economics Letters 111, 43-46.

Lazo, J.K., Waldman, D.M., Morrow, B.H., Thacher, J.A., 2010. Household Evacuation Decision Making and the Benefits of Improved Hurricane Forecasting: Developing a Framework for Assessment. Weather and Forecasting 25, 207-219.

Letson, D., Sutter, D.S., Lazo, J.K., 2007. Economic Value of Hurricane Forecasts: An Overview and Research Needs. Natural Hazards Review 8, 78-86.

Mangham, L.J., Hanson, K., McPake, B., 2009. How to do (or not to do) ... Designing a discrete choice experiment for application in a low-income country. Health Policy Plan. 24, 151-158.

Mazumdar, T., Raj, S.P., Sinha, I., 2005. Reference Price Research: Review and Propositions. Journal of Marketing 69, 84-102.

McNair, B.J., Bennett, J., Hensher, D.A., 2011. A comparison of responses to single and repeated discrete choice questions. Resource and Energy Economics 33, 554-571.

Millner, A., 2008. Getting the Most out of Ensemble Forecasts: A Valuation Model Based on User–Forecast Interactions. Journal of Applied Meteorology and Climatology 47, 2561-2571.

Mjelde, J.W., Peel, D.S., Sonka, S.T., Lamb, P.J., 1993. Characteristics of Climate Forecast Quality: implications for Economic Value to Midwestern Corn Producers. Journal of Climate 6, 2175-2187.

MONRE, 2009. Climate change, sea leve rise scenarios for Vietnam. Ministry of Natural resource and Environment (MONRE), 6/2009, Hanoi, Vietnam.

Morrison, M., Bennett, J., Blamey, R., Louviere, J., 2002. Choice Modeling and Tests of Benefit Transfer. American Journal of Agricultural Economics 84, 161-170.

Mosquera-Machado, S., Dilley, M., 2009. A comparison of selected global disaster risk assessment results. Natural Hazards 48, 439-456.

Nguyen, T.C., Robinson, J., 2013. Analysing motives behind willingness to pay for improving early warning services for tropical cyclones in Vietnam. Meteorological Applications, Early View (DOI: 10.1002/met.1441).

Nguyen, T.C., Robinson, J., Kaneko, S., Komatsu, S., 2013. Estimating the value of economic benefits associated with adaptation to climate change in a developing country: A case study of improvements in tropical cyclone warning services. Ecological Economics 86, 117-128.

Othman, J., Bennett, J., Blamey, R., 2004. Environmental values and resource management options: a choice modelling experience in Malaysia. Environment and Development Economics 9, 803-824.

Peduzzi, P., Chatenoux, B., Dao, H., De Bono, A., Herold, C., Kossin, J., Mouton, F., Nordbeck, O., 2012. Global trends in tropical cyclone risk. Nature Clim. Change 2, 289-294.

PEP, 2008. The Environment, Income generation and the Poor. Poverty and Environment Program of Vietnam Ministry of Natural Resources and Environment (MONRE) and United Nations Development Programme (UNDP), Hanoi, Vietnam.

Poe, G.L., Welsh, M.P., Champ, P.A., 1997. Measuring the Difference in Mean Willingness to Pay When Dichotomous Choice Contingent Valuation Responses Are Not Independent. Land Economics 73, 255-267.

Putler, D.S., 1992. Incorporating Reference Price Effects into a Theory of Consumer Choice. Marketing Science 11, 287-309.

Scheufele, G., Bennett, J., 2012. Response Strategies and Learning in Discrete Choice Experiments. Environmental and Resource Economics 52, 435-453.

Scheufele, G., Bennett, J., 2013. Effects of alternative elicitation formats in discrete choice experiments. Australian Journal of Agricultural and Resource Economics, no-no.

Train, K.E., 2009. Discrete choice methods with simulation, 2nd ed. Cambridge University Press, New York.

Tuan, T., Navrud, S., 2007. Valuing cultural heritage in developing countries: comparing and pooling contingent valuation and choice modelling estimates. Environmental and Resource Economics 38, 51-69.

Whittington, D., 1998. Administering contingent valuation surveys in developing countries. World Development 26, 21-30.

Figure 1: An example of choice task

	Current Situation	Proposed Improvements
Accuracy of tropical cyclone forecast	Current condition	Improvement **LEVEL 2**
Number of **updates** per day	8 times	**16 times**
Mobile phone short message warning	Not available	**Available**
A **one-off payment in your electricity bill**	None	**150,000 VND**

Source: Nguyen et al. (2013)

Figure 2: The provinces (with larger dots) where our DCE surveys were undertaken.

Figure 3: Changes relative to position 1 in acceptance rates of improvement alternatives and of options containing accuracy level 1

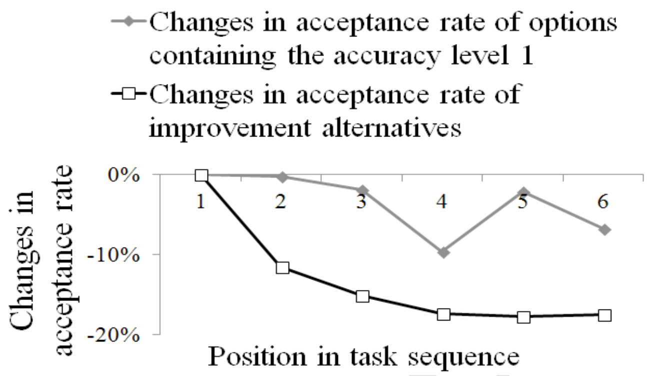


Figure 4: Changes relative to the position 1 in acceptance rates of options containing different levels of accuracy improvement

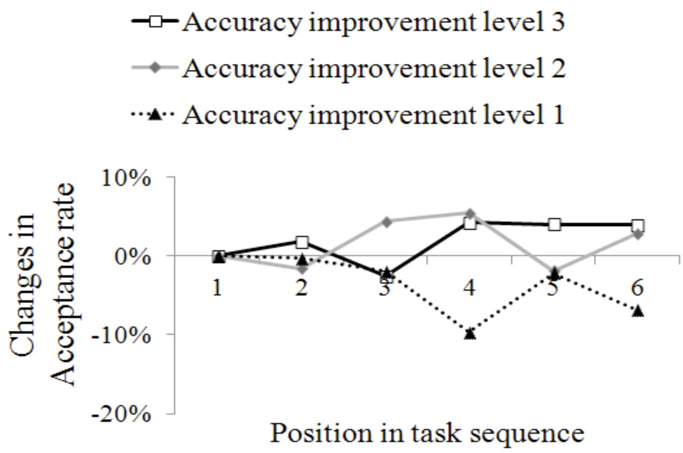


Figure 5: Acceptance rates of options containing different cost levels

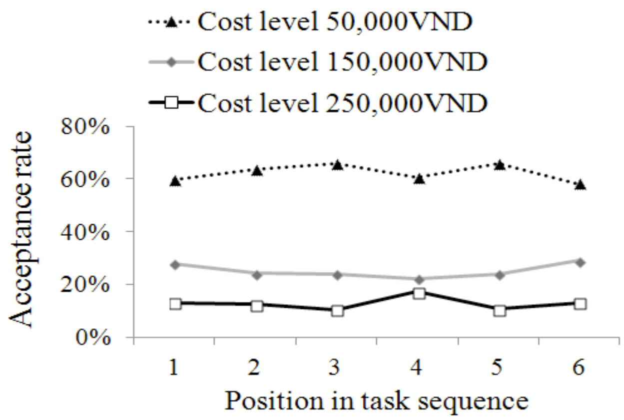


Figure 6: Changes relative to position 1 in acceptance rates of options containing different cost levels

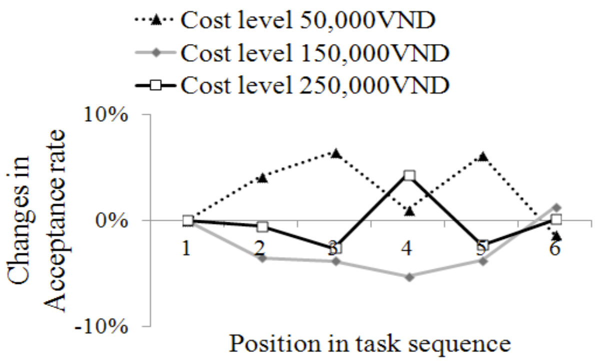
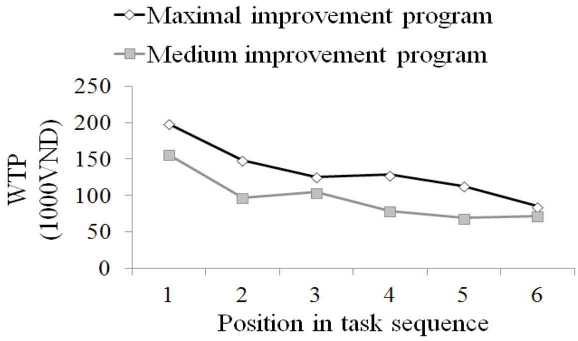
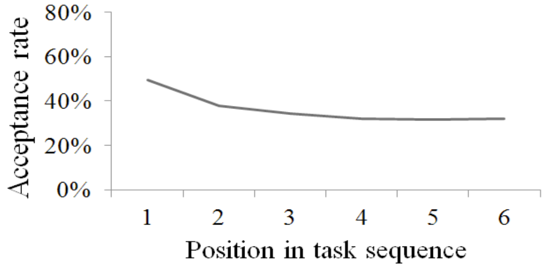


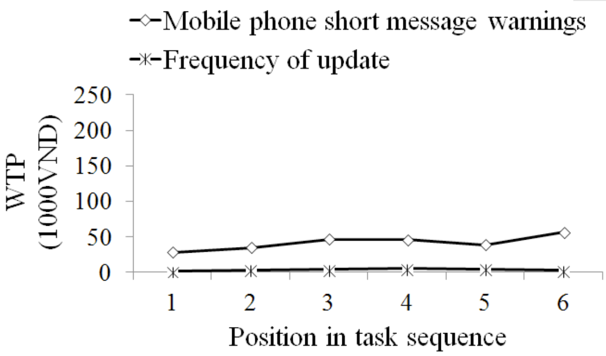
Figure 7: Comparisons between trends in WTP estimates and in acceptance rates across the sequence of choice tasks



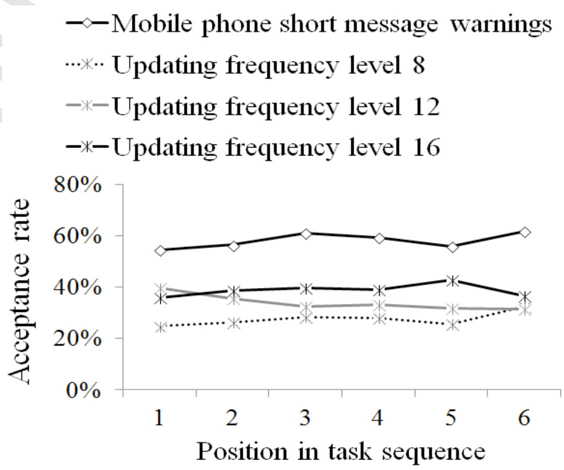
(a1) Total WTP estimates



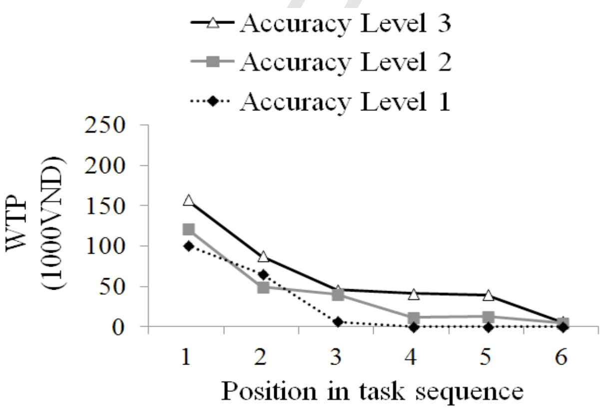
(a2) Acceptance rate of improvement alternatives



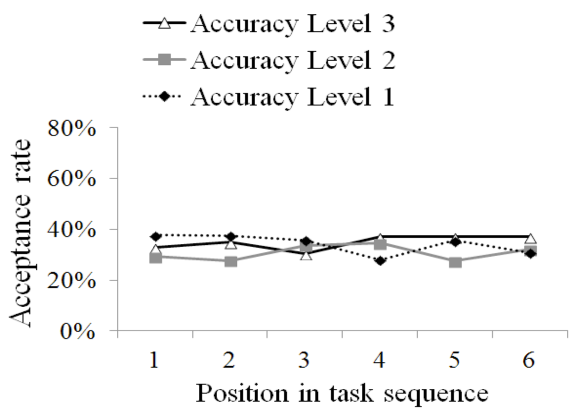
(b1) WTP estimates for mobile phone based warning service and a unit change in frequency of update



(b2) Acceptance rates of options containing mobile phone based warning and different levels of updating frequency



(c1) WTP estimates for the three accuracy improvement levels



(c2) Acceptance rates of options containing different levels of accuracy improvement

Table 1: Attributes and levels for the discrete choice experiment in Vietnam

Attributes	Current levels	Improvement levels
Accuracy of tropical cyclone forecast	Current condition	Level 1, Level 2, Level 3
Number of updates per day (times)	8	8, 12, 16
Mobile phone short message warning	Not available	Not available, Available
A one-off payment in household electricity bill (1000 VND[a])	0	50, 150, 250

[a] 1 USD = 20,800 VND (from website of State Bank of Vietnam, accessed on 21/11/2011)

Source: Nguyen et al. (2013)

Table 2: Information about survey sites

Survey sites	Location characteristics			Number of respondents	Response rate
	Province	Urban or Rural area	Central or North region		
Tam Tien Commune	Quang Nam	Rural	Central	256	93%
Son Tra District	Da Nang	Urban	Central	255	86%
Vinh Quang Commune	Hai Phong	Rural	North	248	96%
Do Son District	Hai Phong	Urban	North	255	86%
Total				1014	89%

Table 3: Socio-economic characteristics of the surveyed sample and Vietnamese population

Socio-economic characteristics	Vietnamese population	The surveyed sample[a]	P-value[b] of tests of differences
Male (% male)	49.4%	83.6% (0.37)	0.000
Age (> = 18 years)	39.6	47.5 (10.19)	0.000
Income (monthly mill.VND per household)	5.4	4.4 (3.32)	0.000
Household size (# persons in household)	3.89	4.22 (1.52)	0.000
Education (% with more than high school)	30.0%	32.3% (0.47)	0.103

[a] Mean values (standard deviations) are reported; [b] P-values are calculated from two-sided t-tests of differences in age, household size and income, and from chi-squared tests of differences in proportions of male respondents and respondents with high school degree.

Table 4: P-values estimated from the Wilcoxon signed-rank tests

	Position 1	Position 2	Position 3	Position 4	Position 5	Position 6
Position 1	-	-	-	-	-	-
Position 2	0.000***	-	-	-	-	-
Position 3	0.000***	0.049	-	-	-	-
Position 4	0.000***	0.018	0.511	-	-	-
Position 5	0.000***	0.001***	0.236	0.927	-	-
Position 6	0.000***	0.004***	0.219	0.886	0.784	-

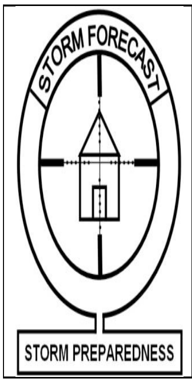
*** denotes 1% significance level

Table 5: ML model with ordering-related variables

Variables	Mean	Standard deviation	Variables	Mean	Standard deviation
ASC	2.623***	2.214***	Frequency of update	0.043	0.060**
	(0.368)	(0 .261)		(0.033)	(0.024)
Position 2	-1.104**		Position 2	0.033	
	(0.518)			(0.056)	
Position 3	-2.489***		Position 3	0.044	
	(0.537)			(0.059)	
Position 4	-3.178***		Position 4	0.049	
	(0.550)			(0.058)	
Position 5	-2.625***		Position 5	0.040	
	(0.549)			(0.059)	
Position 6	-2.985***		Position 6	0.0003	
	(0.571)			(0.057)	
Accuracy level 2	0.548*	1.031***	Mobile phone short	0.743***	2.291***
	(0.291)	(0.151)	message warning	(0.245)	(0.367)
Position 2	-0.927**		Position 2	0.069	
	(0.420)			(0.346)	
Position 3	0.162		Position 3	0.230	
	(0.432)			(0.360)	
Position 4	0.226		Position 4	0.083	
	(0.439)			(0.356)	
Position 5	-0.298		Position 5	0.002	
	(0.428)			(0.358)	
Position 6	-0.116		Position 6	0.133	
	(0.441)			(0.363)	
Accuracy level 3	1.477***	1.810***	Cost	-0.026***	
	(0.341)	(0.268)		(0.002)	
Position 2	-0.977**		Position 2	0.003	
	(0.476)			(0.003)	
Position 3	-0.665		Position 3	0.005*	
	(0.531)			(0.003)	
Position 4	-0.188		Position 4	0.008***	
	(0.517)			(0.003)	
Position 5	-0.718		Position 5	0.007**	
	(0.504)			(0.003)	
Position 6	-1.018**		Position 6	0.011***	
	(0.503)			(0.003)	
			Best deal	0.664	
				(0.440)	
			Worst deal	1.621**	
				(0.648)	
			First deal	1.946***	
Number of respondents	1014			(0.381)	
Log-likelihood	-2706.871		Previous deal	0.178	
Pseudo-R^2	0.358			(0.370)	

Standard deviations are in parentheses; *** denotes 1% significance level, ** denotes 5% significance level, * denotes 10% significance level; ASC: alternative specific constant, equal to 1 for improvement alternatives.

Appendix A: **Description of "Accuracy of tropical cyclone forecast"**



Tropical cyclone forecast contains two main categories of projected information: (1) tropical cyclone position; (2) landfall time

Tropical cyclone position forecast helps to find out "whether the tropical cyclone would move toward your place or not?"

Landfall time forecast helps to find out "when you need to do the preventing activities?"

More accurate forecast information means the smaller ranges of errors (or the dark ranges in the below table). The more exact forecast information could help to not only reduce tropical cyclone damages in the case tropical cyclones affect your place, but also avoid costs of unnecessary preventing activities while tropical cyclones actually don't affect your place.

		Error of TROPICAL CYCLONE POSITION FORECAST	Error of EXPECTED TIME OF LANDFALL
Current condition		140 km – 140 km	Before 6 hours – After 12 hours
Improvement levels	Level 1	100 km – 100 km	Before 4 hours – After 8 hours
	Level 2	80 km – 80 km	Before 3 hours – After 6 hours
	Level 3	60 km – 60 km	Before 2 hours – After 4 hours

Source: Nguyen and Robinson (2013)

Appendix B: Acceptance rates and number of respondents for each choice task across the six positions

	Position 1		Position 2		Position 3		Position 4		Position 5		Position 6	
	Acceptance rate (%)	Number of respondents	Acceptance rate (%)	Number of respondents	Acceptance rate (%)	Number of respondents	Acceptance rate (%)	Number of respondents	Acceptance rate (%)	Number of respondents	Acceptance rate (%)	Number of respondents
Task 1	36	44	33	42	42	53	37	38	31	45	35	31
Task 2	86	59	84	44	68	41	67	39	76	38	66	32
Task 3	19	31	11	45	7	43	15	41	2	41	6	52
Task 4	76	45	77	35	80	44	70	46	62	47	75	36
Task 5	29	35	14	49	9	32	5	43	5	44	10	50
Task 6	28	39	13	38	10	40	9	46	16	38	12	52
Task 7	5	37	5	41	0	43	2	47	2	42	5	43
Task 8	62	53	63	35	54	41	46	35	53	57	31	32
Task 9	48	27	19	47	28	39	16	44	13	38	34	58
Task 10	16	43	10	42	0	56	7	46	0	28	3	38
Task 11	85	47	77	48	80	41	84	32	79	43	76	42
Task 12	70	46	40	40	30	33	39	49	29	45	40	40
Task 13	83	46	78	46	88	33	67	39	75	44	75	44
Task 14	44	48	33	48	27	33	9	34	23	47	26	42
Task 15	29	35	16	38	21	48	29	45	8	38	17	48
Task 16	46	37	26	43	15	46	16	45	23	40	22	41
Task 17	78	49	63	40	49	47	38	42	38	42	44	32
Task 18	27	37	24	37	9	45	19	47	15	41	13	45
Task 19	83	52	72	43	58	40	63	46	57	42	55	33
Task 20	16	31	12	49	12	41	14	44	15	41	16	50
Task 21	40	43	9	43	16	45	9	44	18	39	12	42
Task 22	51	35	30	33	20	46	23	52	24	51	15	39
Task 23	82	51	68	41	73	44	47	38	59	39	56	43
Task 24	50	44	32	47	30	40	44	32	41	44	24	49
Total		1014		1014		1014		1014		1014		1014
Mean	50		38		34		32		32		32	

Appendix C: The confounding between the alternative specific constant and the accuracy variables

In this DCE study, there is a confounding between qualitative parameters of the accuracy attribute and an alternative specific constant (ASC) in the utility function. This confounding is caused by a decision not to include the status quo level of the accuracy attribute in the improvement options. A main reason for this decision is that not including the status quo level into the improved options is expected to make choice options more realistic. In 2010, the Government of Vietnam endorsed the Strategy for Development of Hydro-meteorological Service which is operational until 2020. Under the Strategy, one improvement program ratified on 25/06/2010 by the Prime Minister was to modernise forecasting technologies and monitoring systems for hydro-meteorological services in 2010 – 2012. It is expected that the investment in forecasting technologies and monitoring systems would create improvements in forecast accuracy. The Government's strategy and the improvement program were briefly introduced to respondents at the beginning of the survey to emphasise the survey's creditability to respondents. If no improvement in accuracy was included in the improvement alternatives, respondents would think that no improvement in accuracy was unrealistic, particularly because the Government's improvement program commenced one year before the commencement of the survey. In developing a DCE exercise, it is very important to attempt to construct the choice task with as much realism as possible (Alpízar et al., 2001; Bennett and Adamowicz, 2001). Unrealistic choice options could result in zero-bid protest votes against all proposed program (Boxall et al., 2012). In addition, some respondents, who think the options are unrealistic and do not believe that the survey is credible, may still vote "yes" for some reasons (e.g. to please the interviewers) or may have random answers to not only choice questions but also follow-up questions. In that case, the responses may introduce unidentified biases to the results. However, it is acknowledged that the design of the accuracy attribute in this study causes problems in the econometric analysis of this DCE survey.

- *Approach to modelling the accuracy attribute*

If the accuracy attribute was treated as a quantitative variable, there would not be a confounding problem. However, a body of literature on valuation of meteorological services has indicated that accuracy and forecast value have a non-linear relationship (Letson et al., 2007; Millner, 2008; Mjelde et al., 1993). The accuracy attribute should be measured as a qualitative variable to capture the non-linear relationship between WTP values and the accuracy of cyclone forecasting. The accuracy attribute has four levels, including the status quo level and three improvement levels 1, 2, and 3. These four levels require three accuracy dummy parameters: accuracy level 1, accuracy level 2 and accuracy level 3. Since the accuracy parameters are confounded with an ASC, the dummy parameter representing the preference associated with accuracy level 1 is excluded in order to estimate econometric models in this study. In the results of econometric models, the ASC is reported with the implicit assumption that the mean coefficient of accuracy level 1 is 0. With the exclusion of the accuracy level 1 parameter, the ASC captures the preference for accuracy level 1, so that the ASC represents both respondents' utility of moving away from the current situation and their preference for accuracy improvement level 1. The approach to modelling the accuracy attribute used in this study is similar to the approach used in the Boxall et al. (2012) study, where there is a similar situation that a constant is confounded with the least improved program (program A).

- *Estimating total willingness-to-pay*

This section continues with a discussion on estimating total WTP, since an understanding of how total WTP was estimated in this research will clarify the interpretation of WTP estimates for the accuracy improvement levels. Applying the previously discussed approach to modelling the accuracy attribute, a conditional logit (CL) model was estimated in order to present more detailed discussion on how the WTP estimates were calculated in this research. The CL model is specified as follows:

Alternative 1 (the status quo):

$V_1 = \quad \beta_1$ * Accuracy Level 2 + β_2 *Accuracy Level 3 + β_3 * Frequency of update + β_4 * Mobile phone short message warning + β_5 * Cost

Alternative 2 (proposed improvements):

V_2 = ASC + β_1 * Accuracy Level 2 + β_2 *Accuracy Level 3 + β_3 * Frequency of update + β_4 * Mobile phone short message warning + β_5 * Cost

In the CL model, the ASC is equal to 1 for improvement alternatives, and 0 for the status quo option. Accuracy levels 2 and 3, which are dummy variables, are assessed relative to accuracy level 1; and after accounting for the ASC, the implied coefficient of accuracy level 1 is 0. Frequency of update and cost variables are treated as continuous variables, while mobile phone short message warning is modelled as a dummy variable. The parameter estimates are contained in Table C.1.

Table C.1: CL model results

Variable	Coefficient name	Mean estimates
ASC	ASC	0.493*** (0.092)
Accuracy level 2	β_1	0.303*** (0.077)
Accuracy level 3	β_2	0.907*** (0.082)
Frequency of update	β_3	0.060*** (0.010)
Mobile phone short message warning	β_4	0.692*** (0.032)
Cost	β_5	-0.015*** (0.0004)
Log-likelihood		-3172.295
Pseudo-R^2		0.248
Number of observations		6084

Standard errors are in parentheses; *** denotes 1% significance level.

When one improvement program is assessed relative to the current situation, compensating surplus (CS) welfare estimates can be obtained in the following formula (Boxall et al., 1996):

$$CS = -\frac{1}{\alpha}(V_{0n} - V_{1n})$$

where α is the marginal utility of income (represented by the β coefficient of the cost attribute), and V_{1n} and V_{0n} are indirect utility functions with and without a specified change in the non-market good or service, respectively. Total WTP estimates are reported in Table C.2. The total WTP is estimated for two improvement programs described as follows:

+ Medium improvement: accuracy improvement level 2, update frequency of 12 times and mobile phone short message warning.

+ Maximal improvement: accuracy improvement level 3, update frequency of 16 times and mobile phone short message warning.

Table C.2. Total WTP estimates from the CL model (1000VND)

Program	Formula	Mean estimate
Medium improvement	$-(ASC+ \beta_1 + 4\times\beta_3+\beta_4)/ \beta_5$	115.310*** (4.024)
Maximal improvement	$-(ASC+ \beta_2 + 8\times\beta_3+\beta_4)/ \beta_5$	171.675*** (4.954)

Standard errors are in parentheses, and are based on the Krinksy–Robb simulation using 1000 draws;
*** denotes 1% significance level

As can be seen in Table C.2, the total WTP estimates include the preferences associated with the ASC. The question arises as to whether it is acceptable to include the ASC into total WTP calculations. When including the ASC, total WTP values for an improvement program are adjusted by the amount of utility associated with moving away from the status quo. There is not clear guidance in the literature on inclusion or exclusion of the preferences expressed in the ASC for the calculation of total WTP (Adamowicz et al., 2011). A number of previous studies included the ASC, since it may represent unobserved valid preferences (Adamowicz et al., 1998; Adamowicz et al., 2011; Blamey et al., 2000; Morrison et al., 2002). Adamowicz et al. (1998) suggest that the total WTP measure, including the ASC, "should probably be considered the most accurate measure" (page 73). However, there is concern that the preference associated with moving away from the status quo may be vulnerable to "yea-saying" responses. With the collaboration with Vietnamese government agencies, it is expected that respondents could believe the survey as being “consequential” with real policy implications and monetary repercussions; such that it might help to minimise "yea-saying" bias. Given treatments to "yea-saying" bias, it is believed that the inclusion of the preference expressed in the ASC could result in more accurate estimates of total WTP. When the ASC is included in the total WTP estimation, the confounding between the preference associated with the ASC and the WTP for accuracy level 1 would not affect the total WTP estimates in this research.

- *Estimating willingness-to-pay for the attributes*

Due to the confounding between the preference for accuracy level 1 and the utility associated with moving away from the status quo, the coefficient of the ASC is used to estimate WTP for the accuracy improvement levels. When the accuracy level 1 parameter is excluded, the ASC reported in Table 1 represents the sum of the utility associated with moving away from the status quo and the preference for accuracy improvement level 1. With only one piece of information on the coefficient of the ASC, one of the following two assumptions must be chosen in order to estimate WTP for the accuracy improvement levels: Assumption 1 - the WTP for accuracy level 1 = 0; Assumption 2 - the utility associated with moving away from the status quo = 0, so that the WTP for the accuracy levels can be estimated using the coefficient of the ASC. Assumption 1, in which WTP for accuracy improvement level 1 is 0, is not relevant to the research objective of providing WTP estimates for all attributes. Assumption 2, therefore, is applied for the WTP estimation for the accuracy improvement levels in this research.

With assumption 2, the ASC coefficient could be used to estimate WTP for accuracy level 1. Accuracy improvement levels 2 and 3 are modelled relative to accuracy level 1. The WTP estimates for improvements from the status quo level to accuracy levels 2 and 3 are estimated by the sum of the WTP for accuracy level 1 and WTP for improvements in accuracy from level 1 to level 2 and level 3, respectively. The formulas for calculating the WTP estimates for the accuracy improvement levels in this research are presented in Table C.3.

Table C.3: WTP estimates for the attributes (1000VND)

Attribute	Formula	Mean estimate
Accuracy Level 1	$-ASC/\beta_5$	32.903*** (5.991)
Accuracy Level 2	$-(ASC+\beta_1)/\beta_5$	53.111*** (5.353)
Accuracy Level 3	$-(ASC+\beta_2)/\beta_5$	93.466*** (5.379)
Frequency of update	$-\beta_3/\beta_5$	4.002*** (0.697)
Mobile phone short message warning	$-\beta_4/\beta_5$	46.190*** (4.333)

Standard errors are in parentheses, and are based on the Krinksy–Robb simulation using 1000 draws;
*** denotes 1% significance level.

Assumption 2, in which the utility associated with moving away from the status quo is 0, could be criticised as unrealistic. As discussed above, the estimation of total WTP includes the utility associated with moving away from the status quo (i.e. the amount of utility expressed in the ASC). To relax the assumption that the utility associated with moving away from the status quo is 0, the WTP estimates for accuracy level 1, level 2 and level 3 should be interpreted as total WTP estimates for a program with only the accuracy attribute improved up to level 1, level 2 and level 3, respectively. With the interpretation as total WTP estimates, the WTP estimates for the accuracy improvement levels in this research would still provide meaningful information.

Appendix D: P-values of resampling tests for difference in WTP estimates

	Position 1	Position 2	Position 3	Position 4	Position 5	Position 6
Total WTP for medium improvement program						
Position 1	-	-	-	-	-	-
Position 2	0.000***	-	-	-	-	-
Position 3	0.000***	0.340	-	-	-	-
Position 4	0.000***	0.195	0.108	-	-	-
Position 5	0.000***	0.084	0.036	0.316	-	-
Position 6	0.001***	0.171	0.111	0.388	0.451	-
Total WTP for maximal improvement program						
Position 1	-	-	-	-	-	-
Position 2	0.004***	-	-	-	-	-
Position 3	0.001***	0.173	-	-	-	-
Position 4	0.003***	0.224	0.463	-	-	-
Position 5	0.000***	0.072	0.310	0.284	-	-
Position 6	0.000***	0.023	0.096	0.084	0.168	-
Marginal WTP for accuracy improvement level 1						
Position 1	-	-	-	-	-	-
Position 2	0.022	-	-	-	-	-
Position 3	0.000***	0.008***	-	-	-	-
Position 4	0.000***	0.000***	0.115	-	-	-
Position 5	0.000***	0.006***	0.408	0.158	-	-
Position 6	0.000***	0.005***	0.200	0.418	0.257	-
Marginal WTP for accuracy improvement level 2						
Position 1	-	-	-	-	-	-
Position 2	0.000***	-	-	-	-	-
Position 3	0.000***	0.349	-	-	-	-
Position 4	0.000***	0.076	0.139	-	-	-
Position 5	0.000***	0.085	0.149	0.488	-	-
Position 6	0.000***	0.087	0.136	0.402	0.397	-
Marginal WTP for accuracy improvement level 3						
Position 1	-	-	-	-	-	-
Position 2	0.000***	-	-	-	-	-
Position 3	0.000***	0.050	-	-	-	-
Position 4	0.000***	0.062	0.443	-	-	-
Position 5	0.000***	0.044	0.421	0.482	-	-
Position 6	0.000***	0.015	0.150	0.175	0.178	-

	Position 1	Position 2	Position 3	Position 4	Position 5	Position 6
Marginal WTP for frequency of update						
Position 1	-	-	-	-	-	-
Position 2	0.239	-	-	-	-	-
Position 3	0.166	0.366	-	-	-	-
Position 4	0.110	0.262	0.382	-	-	-
Position 5	0.173	0.362	0.483	0.402	-	-
Position 6	0.355	0.441	0.343	0.255	0.334	-
Marginal WTP for mobile phone short message warning						
Position 1	-	-	-	-	-	-
Position 2	0.324	-	-	-	-	-
Position 3	0.122	0.238	-	-	-	-
Position 4	0.155	0.259	0.485	-	-	-
Position 5	0.257	0.406	0.336	0.353	-	-
Position 6	0.084	0.151	0.334	0.328	0.220	-

*** denotes 1% significance level